

RANDOM WALK ON UNIPOTENT MATRIX GROUPS

PERSI DIACONIS AND BOB HOUGH

ABSTRACT. We introduce a new method for proving central limit theorems for random walk on nilpotent groups. The method is illustrated in a local central limit theorem on the Heisenberg group, weakening the necessary conditions on the driving measure. As a second illustration, the method is used to study walks on the $n \times n$ uni-upper triangular group with entries taken modulo p . The method allows sharp answers to the behavior of individual coordinates: coordinates immediately above the diagonal require order p^2 steps for randomness, coordinates on the second diagonal require order p steps; coordinates on the k th diagonal require order $p^{\frac{2}{k}}$ steps.

1. INTRODUCTION

Let $\mathbb{H}(\mathbb{R})$ denote the real Heisenberg group

$$\mathbb{H}(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

Abbreviate $\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$ with $[x, y, z]$, identified with a vector in \mathbb{R}^3 .

Consider simple random walk on $G = \mathbb{H}(\mathbb{R})$ driven by Borel probability measure μ . For $N \geq 1$, the law of this walk is the convolution power μ^{*N} where, for measures ν, ξ on G , and for $f \in C_c(G)$,

$$\langle f, \nu * \xi \rangle = \int_{g, h \in G} f(gh) \, d\nu(g) \, d\xi(h).$$

Say that measure μ is non-lattice (aperiodic) if its support is not contained in a proper closed subgroup of G . For general non-lattice μ of compact support Breuillard [6] uses the representation theory of G to prove a local limit theorem for the law of μ^{*N} , asymptotically evaluating its density in translates of bounded Borel sets. However, in evaluating μ^{*N} on Borel sets translated on both the left and the

2010 *Mathematics Subject Classification.* Primary 60F05, 60B15, 20B25, 22E25, 60J10, 60E10, 60F25, 60G42.

Key words and phrases. Random walk on a group, Heisenberg group, local limit theorem, unipotent group.

We are grateful to Laurent Saloff-Coste, who provided us a detailed account of previous work.

right he makes a decay assumption on the Fourier transform of the abelianization of the measure μ , and raises the question of whether this is needed. We show that this condition is unnecessary. In doing so we give an alternative approach to the local limit theorem on G treating it as an extension of the classical local limit theorem on \mathbb{R}^n , with the further advantage that our argument applies without significant change to arbitrary μ of compact support. We also obtain the optimal rate. The method of argument is related to, though simpler than, the analysis of quantitative equidistribution of polynomial orbits on G from [18].

Recall that the abelianization $G_{\text{ab}} = G/[G, G]$ of G is isomorphic to \mathbb{R}^2 with projection $p : G \rightarrow G_{\text{ab}}$ given by $p([x, y, z]) = [x, y]$. Assume that the probability measure μ satisfies the following conditions.

- (1) *Compact support.*
- (2) *Centered.* The projection p satisfies

$$\int_G p(g) d\mu(g) = 0.$$

- (3) *Lazy.* For all open sets N with $\text{id} \in N$, $\mu(N) > 0$.
- (4) *Full dimension.* Let $\Gamma = \overline{\langle \text{supp } \mu \rangle}$ be the closure of the subgroup of G generated by the support of μ . The quotient G/Γ is compact.

Section 2 gives a characterization of closed subgroups Γ of G of full dimension. Note that a closed subgroup Γ_{ab} of full dimension in the abelianization is isomorphic to one of \mathbb{R}^2 , $\mathbb{R} \times \mathbb{Z}$, \mathbb{Z}^2 . In the case of \mathbb{Z}^2 , there are $v_1, v_2 \in \mathbb{R}^2$ such that $\Gamma_{\text{ab}} = \mathbb{Z}v_1 + \mathbb{Z}v_2$; set $V = |v_1 \wedge v_2|$ for the volume of the lattice, and define the parity function $\varepsilon : \Gamma \rightarrow \mathbb{Z}/2\mathbb{Z}$ by $\varepsilon(n_1v_1 + n_2v_2) \equiv n_1n_2 \pmod{2}$. In either case $\mathbb{R} \times \mathbb{Z}$ or \mathbb{Z}^2 , write $\Gamma_{\text{ab, nl}}$ for the non-lattice component of Γ_{ab} .

Under the above conditions, the central limit theorem for μ is known, see the history at the end of the introduction. Let $(d_t)_{t>0}$ denote the semigroup of dilations given by $d_t([x, y, z]) = [tx, ty, t^2z]$ and denote the Gaussian semigroup $(\nu_t)_{t>0}$ defined by its generator (see [6], [10])

$$\begin{aligned} \mathcal{A}f &= \left. \frac{d}{dt} \right|_{t=0} \int_{g \in G} f(g) d\nu_t(g) \\ &= \bar{z} \partial_z f(\text{id}) + \overline{xy} \partial_{xy}^2 f(\text{id}) + \frac{1}{2} \overline{x^2} \partial_x^2 f(\text{id}) + \frac{1}{2} \overline{y^2} \partial_y^2 f(\text{id}) \end{aligned}$$

where $\sigma_x^2 = \overline{x^2} = \int_{g=[x,y,z] \in G} x^2 d\mu(g)$ and similarly $\sigma_y^2 = \overline{y^2}$, $\sigma_{xy}^2 = \overline{xy}$, \bar{z} . With $\nu = \nu_1$, the central limit theorem for μ states that for $f \in C_c(G)$,

$$\left\langle f, d_{\frac{1}{\sqrt{N}}} \mu^{*N} \right\rangle \rightarrow \langle f, \nu \rangle.$$

For $g \in G$ define the left and right translation operators $L_g, R_g : L^2(G) \rightarrow L^2(G)$,

$$L_g f(h) = f(gh), \quad R_g f(h) = f(hg).$$

Our local limit theorem is as follows.

Theorem 1. *Let μ be a Borel probability measure of compact support on $G = \mathbb{H}(\mathbb{R})$, which is centered, lazy and full dimension. Let, as above, $\Gamma = \overline{\langle \text{supp } \mu \rangle}$ be the closure of the group generated by the support of μ , let \mathcal{F} be a fundamental domain for G/Γ having volume $M > 0$ and set $\chi_{\mathcal{F}} = \frac{1}{M}$ for its unit density. Let ν be the limiting Gaussian measure of $d_{\frac{1}{\sqrt{N}}} \mu^{*N}$. Uniformly for $g, h \in G$ and for $f \in C_c(G)$, for $N \geq 1$,*

$$(1) \quad \langle L_g R_h f, \chi_{\mathcal{F}} * \mu^{*N} \rangle = \langle L_g R_h f, d_{\sqrt{N}} \nu \rangle + o_{\mu}(\|f\|_1 N^{-2}).$$

If any of the following conditions holds

- i. Γ is a discrete subgroup of $\mathbb{H}(\mathbb{R})$
- ii. Γ is not discrete, but Γ_{ab} is a discrete subgroup of \mathbb{R}^2 and the Cramér condition holds

$$\sup_{\substack{\alpha, \beta \in \mathbb{R} \\ a \in \{0,1\} \\ |\lambda| > 1}} \left| \int_{g=[x,y,z] \in G} e^{-i(\alpha x + \beta y + \frac{2\pi\lambda}{V}(z - \frac{xy}{2} + \frac{a\varepsilon(x,y)V}{2}))} d\mu(g) \right| < 1$$

- iii. Γ_{ab} is not discrete, but the Cramér condition holds:

$$\sup_{\lambda \in \widehat{\Gamma_{\text{ab}, \text{nl}}}, |\lambda| > 1} \left| \int_{g=[x,y,z] \in G} e^{-i\lambda \cdot (x,y)} d\mu(g) \right| < 1$$

then uniformly for $g, h \in G$ and $f \in C_c(G)$,

$$(2) \quad \langle L_g R_h f, \chi_{\mathcal{F}} * \mu^{*N} \rangle = \langle L_g R_h f, d_{\sqrt{N}} \nu \rangle + O_{f,\mu}(N^{-\frac{5}{2}}).$$

Remark. The rate is best possible as may be seen by projecting to the abelianization. A variety of other statements of the local theorem are also derived, see eqn. (9) in Section 3.

Remark. For non-lattice μ , [6] obtains (1) with $h = \text{id}$ and for general h subject to Cramér's condition. A condition somewhat weaker than Cramér's would suffice to obtain (2) in the case of ii and iii.

Remark. In the case that μ is discrete or has a density, [1, 2] obtains an error of $O(N^{-\frac{5}{2}})$ in approximating $\mu^{*N}(g)$, $g \in \Gamma$ with the corresponding heat kernel.

The estimates of Theorem 1 permit the following asymptotic evaluation of the probability of return to the identity in simple random walk on $\mathbb{H}(\mathbb{Z})$.

Corollary 2. *Let μ_0 be the measure on $\mathbb{H}(\mathbb{Z})$ which assigns equal probability $\frac{1}{5}$ to the each element of the generating set*

$$\left\{ e, \begin{pmatrix} 1 & \pm 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \pm 1 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

As $N \rightarrow \infty$, $\mu_0^{*N}(e) = \frac{25}{16N^2} + O\left(N^{-\frac{5}{2}}\right)$.

The basic idea which drives the proof of Theorem 1 is that permuting segments of generators in a typical word of the walk generates smoothness in the central coordinate of the product, while leaving the abelianized coordinates unchanged. This observation permits passing from a limit theorem to a local limit theorem by smoothing at a decreasing sequence of scales. As a further application of this technique, answering a question of [13] we determine the mixing time of the central coordinate in a natural class of random walks on the group $N_n(\mathbb{Z}/p\mathbb{Z})$ of $n \times n$ uni-upper triangular matrices with entries in $\mathbb{Z}/p\mathbb{Z}$.

Theorem 3. *Let $n \geq 2$ and let μ be a probability measure on \mathbb{Z}^{n-1} which satisfies the following conditions.*

- (1) Bounded support.
- (2) Full support. $\langle \text{supp } \mu \rangle = \mathbb{Z}^{n-1}$
- (3) Lazy. $\mu(0) > 0$
- (4) Mean zero. $\sum_{x \in \mathbb{Z}^{n-1}} x \mu(x) = 0$
- (5) Trivial covariance.

$$\left(\sum_{x \in \mathbb{Z}^{n-1}} x^{(i)} x^{(j)} \mu(x) \right)_{i,j=1}^{n-1} = I_{n-1}.$$

Push forward μ to a probability measure $\tilde{\mu}$ on $N_n(\mathbb{Z})$ via, for all $x \in \mathbb{Z}^{n-1}$,

$$\tilde{\mu} \left(\begin{pmatrix} 1 & x^{(1)} & 0 & \cdots & 0 \\ 0 & 1 & x^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ & & 0 & 1 & x^{(n-1)} \\ 0 & \cdots & & 0 & 1 \end{pmatrix} \right) = \mu(x).$$

Write $Z : N_n(\mathbb{Z}) \rightarrow \mathbb{Z}$ for the upper right corner entry of a matrix of $N_n(\mathbb{Z})$. There exists $C > 0$ such that, for all primes p , for $N \geq 1$,

$$\sum_{x \bmod p} \left| \tilde{\mu}^{*N}(Z \equiv x \bmod p) - \frac{1}{p} \right| \ll \exp \left(-C \frac{N}{p^{\frac{2}{n-1}}} \right).$$

Remark. Informally, the top right corner entry mixes in time $O\left(p^{\frac{2}{n-1}}\right)$. This is tight, since archimedean considerations show that the L^1 distance to uniform is $\gg 1$ if the number of steps of the walk is $\ll p^{\frac{2}{n-1}}$.

Remark. Although we have considered only the top right corner entry in $U_n(\mathbb{Z}/p\mathbb{Z})$, this result determines the mixing time of each entry above the diagonal by iteratively projecting to the subgroups determined by the top left or bottom right $m \times m$ sub-matrices.

Remark. The argument presented here appears to be sufficient to give a joint local limit theorem for random walk on U_n , $n > 3$ and other nilpotent groups analogous to Theorem 1 for random walk on U_3 , but we have not checked this carefully.

HISTORY

Random walk on groups is a mature subject with myriad projections into probability, analysis and applications. Useful overviews with extensive references are in [5], [24]. Central limit theorems for random walk on Lie groups were first proved by [27] with [26] carrying out the details for the Heisenberg group. Best possible results under a second moment condition for nilpotent Lie groups are in [23].

A general local limit theorem for the Heisenberg group appears in [6], which contains a useful historical review. There similar conditions to those of our Theorem 1 are made, but the argument treats only the non-lattice case and needs a stronger condition on the characteristic function of the measure projected to the abelianization. Remarkable local limit theorems are in [1, 2]. The setting is groups of polynomial growth, and so “essentially” nilpotent Lie groups via Gromov’s Theorem. The first paper gives quite complete results assuming that the generating measure has a density. The second paper gives results for measures supported on a lattice. The arguments in [2] have been adapted in [7] to give a local limit theorem for non-lattice measures supported on finitely many points.

Just as for the classical abelian case, many variations have been studied. Central limit theorems for walks satisfying a Lindeberg condition on general Lie groups are proved in [25], see also references therein. Large deviations for walks on nilpotent groups are proved in [3]. Central limit theorems on covering graphs with nilpotent automorphism groups are treated in [20, 21]. This allows walks on Cayley graphs with some edges and vertices added and deleted. Brownian motion and heat kernel estimates are also relevant, see [19, 17].

Random walk on finite nilpotent groups are a more recent object of study. Diaconis and Saloff-Coste [15, 14, 13] show that for simple symmetric random walk on $\mathbb{Z}/n\mathbb{Z}$, order n^2 steps are necessary and sufficient for convergence to uniform. The first paper uses Nash inequalities, the second lifts to random walk on the free nilpotent group and applies central limit theorems of Hebisch, Saloff-Coste and finally Harnack inequalities to transfer back to the finite setting. The third paper uses geometric ideas of moderate growth to show that for groups of polynomial growth, diameter-squared steps are necessary and sufficient to reach uniformity. This paper raises the question of the behavior of the individual coordinates on $U_n(\mathbb{Z}/p\mathbb{Z})$ which is finally answered in Theorem 3. A direct non-commuting Fourier approach to $\mathbb{H}(\mathbb{Z}/p\mathbb{Z})$ is carried out in [8], where it is shown that order $p \log p$ steps suffice to

make the central coordinate random, improved here to order p steps, which is best possible. For a review of the $\mathbb{H}(\mathbb{Z})$ results, see [12]. Finally there have been quite a number of papers studying the walk on $U_n(\mathbb{Z}/p\mathbb{Z})$ when both p and n grow. We refer to [22], which contains a careful review and definitive results.

NOTATION AND CONVENTIONS

Vectors from \mathbb{R}^d , $d \geq 1$ are written in plain text w , their coordinates with superscripts $w^{(i)}$, and sequences of vectors with an underline \underline{w} . The sum of a sequence of vectors \underline{w} is indicated \overline{w} . w^t denotes the transpose of w . We have been cavalier in our use of transpose; interpretation of vectors as rows or columns should be clear from the context. We frequently identify matrix elements in the group U_n with vectors from Euclidean space, and have attempted to indicate the way in which the vectors should be interpreted. As a rule of thumb, when the group law is written multiplicatively, the product is in the group U_n , and when additively, in Euclidean space.

The arguments presented use permutation group actions on sequences of vectors. Given integer $N \geq 1$, denote \mathfrak{S}_N the symmetric group on $[N] = \mathbb{Z} \cap [1, N]$, which acts on length N sequence of vectors by permuting the indices:

$$\mathfrak{S}_N \ni \sigma : (w_1, \dots, w_N) \mapsto (w_{\sigma(1)}, \dots, w_{\sigma(N)}).$$

C_2 is the two-element group. For $d \geq 1$, identify C_2^d with the d -dimensional hypercube $\{0, 1\}^d$. $\mathbf{1}_d$ is the element of C_2^d corresponding to the sequence of all 1's on the hypercube. C_2^d acts on sequences of vectors of length 2^d with the j th factor determining the relative order of the first and second blocks of 2^{j-1} elements. To illustrate the action of C_2^2 on $\underline{x} = x_1x_2x_3x_4$:

$$(0, 0) \cdot \underline{x} = x_1x_2x_3x_4$$

$$(1, 0) \cdot \underline{x} = x_2x_1x_3x_4$$

$$(0, 1) \cdot \underline{x} = x_3x_4x_1x_2$$

$$(1, 1) \cdot \underline{x} = x_3x_4x_2x_1.$$

The 2-norm on \mathbb{R}^d is indicated $\|\cdot\|$ and $\|\cdot\|_{(\mathbb{R}/\mathbb{Z})^d}$ denotes distance to the nearest integer lattice point. Given $\xi \in \mathbb{R}^d$, $e_\xi(\cdot)$ denotes the character of \mathbb{R}^d , $e_\xi(x) = e^{2\pi i \xi \cdot x}$. Given $m > 1$ and $a \bmod m$, $e_{a,m}(\cdot)$ denotes the character of $\mathbb{Z}/m\mathbb{Z}$, $e_{a,m}(x) = e^{\frac{2\pi i ax}{m}}$.

Use δ_x to indicate the Dirac delta measure at $x \in \mathbb{R}^d$. Given $f \in C_c(\mathbb{R}^d)$ and measure μ , $\langle f, \mu \rangle$ denotes the bilinear pairing

$$\langle f, \mu \rangle = \int_{\mathbb{R}^d} f(x) d\mu(x).$$

Denote the Fourier transform of function f , resp. the characteristic function of measure μ by, for $\xi \in \mathbb{R}^d$,

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e_{-\xi}(x) f(x) dx, \quad \hat{\mu}(\xi) = \int_{\mathbb{R}^d} e_{-\xi}(x) d\mu(x).$$

For $x \in \mathbb{R}^d$, $T_x f$ denotes function f translated by x ,

$$T_x f(y) = f(y - x), \quad \widehat{T_x f}(\xi) = e_{-\xi}(x) \hat{f}(\xi)$$

and for real $t > 0$, f_t denotes f dilated by t ,

$$f_t(x) = \frac{1}{t} f\left(\frac{x}{t}\right), \quad \hat{f}_t(\xi) = \hat{f}(t\xi).$$

For smooth f , the Plancherel identity is

$$\langle f, \mu \rangle = \int_{\mathbb{R}^d} \hat{f}(\xi) \overline{\hat{\mu}(\xi)} d\xi.$$

For $r \in \mathbb{R}$ and $\sigma > 0$, $\eta(r, \sigma)$ denotes the one-dimensional Gaussian distribution with mean r and variance σ^2 , with density and characteristic function

$$\eta(r, \sigma)(x) = \frac{\exp\left(-\frac{(x-r)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma}, \quad \widehat{\eta(r, \sigma)}(\xi) = e_{-\xi}(r) \exp(-2\pi^2\sigma^2\xi^2).$$

A centered (mean zero) normal distribution η in dimension d is specified by its covariance matrix

$$\underline{\sigma}^2 = \left(\int_{\mathbb{R}^d} x^{(m)} x^{(n)} \eta(x) \right)_{m,n=1}^d$$

and has density and characteristic function

$$\eta(0, \underline{\sigma})(x) = \frac{\exp\left(-\frac{x^t(\underline{\sigma}^2)^{-1}x}{2}\right)}{(2\pi)^{\frac{d}{2}} (\det \underline{\sigma}^2)^{\frac{1}{2}}}, \quad \widehat{\eta(0, \underline{\sigma})}(\xi) = \exp(-2\pi^2 \xi^t \underline{\sigma}^2 \xi).$$

All of our arguments concern the repeated convolution μ^{*N} of a fixed measure μ on the upper triangular matrices. Asymptotic statements are with respect to N as the large parameter. The Vinogradov notation $A \ll B$, resp. $A \gg B$, means $A = O(B)$, resp. $B = O(A)$. $A \asymp B$ means $A \ll B$ and $B \ll A$.

2. BACKGROUND TO THEOREM 1

This section collects together several background statements regarding the Heisenberg group, its Gaussian semigroups of probability measures and statements of elementary probability which are needed in the course of the argument.

Write $A = [1, 0, 0]$, $B = [0, 1, 0]$, $C = [0, 0, 1]$. The following commutators are useful,

$$\begin{aligned} [A, B] &= ABA^{-1}B^{-1} = [0, 0, 1] = C, \\ [A^{-1}, B^{-1}] &= A^{-1}B^{-1}AB = [0, 0, 1] = C, \\ [A, B^{-1}] &= AB^{-1}A^{-1}B = [0, 0, -1] = C^{-1}, \\ [A^{-1}, B] &= A^{-1}BAB^{-1} = [0, 0, -1] = C^{-1}. \end{aligned}$$

A convenient representation for $[x, y, z] \in \mathbb{H}(\mathbb{R})$ is $C^z B^y A^x$. Using the commutator rules above, the multiplication rule for $\underline{w} \in \mathbb{H}(\mathbb{R})^N$ becomes

$$(3) \quad \prod_{i=1}^N [w_i^{(1)}, w_i^{(2)}, w_i^{(3)}] = [\underline{w}^{(1)}, \underline{w}^{(2)}, \underline{w}^{(3)} + H(\underline{w})]$$

where $\bar{\cdot}$ and H act on sequences of vectors from \mathbb{R}^d ($d \geq 1$, resp. $d \geq 2$) via

$$(4) \quad \underline{w} = \sum_i w_i, \quad H(\underline{w}) = \sum_{i < j} w_i^{(1)} w_j^{(2)}.$$

It is also convenient to define

$$\begin{aligned} (5) \quad H^*(\underline{w}) &= H(\underline{w}) - \frac{1}{2} \underline{w}^{(1)} \underline{w}^{(2)} + \frac{1}{2} \sum_{i=1}^N w_i^{(1)} w_i^{(2)} \\ &= \frac{1}{2} \sum_{1 \leq i < j \leq N} \left(w_i^{(1)} w_j^{(2)} - w_i^{(2)} w_j^{(1)} \right), \end{aligned}$$

and for $w = [x, y, z]$, $\tilde{w} = [x, y, z - \frac{1}{2}xy]$, so that the multiplication rule may be written

$$(6) \quad \prod_{i=1}^N w_i = \underline{\tilde{w}} + \left[0, 0, \frac{1}{2} \underline{\tilde{w}}^{(1)} \underline{\tilde{w}}^{(2)} + H^*(\underline{\tilde{w}}) \right].$$

Let $S = \text{supp } \mu$. Recall that $\Gamma = \overline{\langle S \rangle}$ is the closure of the group generated by S . Its abelianization, $\Gamma_{\text{ab}} = \Gamma/[\Gamma, \Gamma]$ is equal to $p(\Gamma)$ where p is the projection $p : G \rightarrow G_{\text{ab}}$. Let Γ_0 be the semigroup generated by S . We record the following descriptions of Γ and Γ_0 .

Proposition 4. *Let $\Gamma \leq \mathbb{H}(\mathbb{R})$ be a closed subgroup of full dimension. The structure of the abelianization $\Gamma_{\text{ab}} = \Gamma/[\Gamma, \Gamma]$ and of Γ falls into one of the following cases.*

(1)

$$\Gamma_{\text{ab}} = \mathbb{R}^2, \quad \Gamma = \{[\gamma, r] : \gamma \in \Gamma_{\text{ab}}, r \in \mathbb{R}\}$$

(2) *There exist non-zero orthogonal $v_1, v_2 \in \mathbb{R}^2$, such that*

$$\Gamma_{\text{ab}} = \{nv_1 + rv_2 : n \in \mathbb{Z}, r \in \mathbb{R}\}, \quad \Gamma = \{[\gamma, r] : \gamma \in \Gamma_{\text{ab}}, r \in \mathbb{R}\}$$

- (3) *There exist non-zero $v_1, v_2 \in \mathbb{R}^2$, linearly independent over \mathbb{R} , such that*

$$\Gamma_{\text{ab}} = \{n_1 v_1 + n_2 v_2 : n_1, n_2 \in \mathbb{Z}\}.$$

In this case, Γ falls into one of two further cases

- (a) $\Gamma = \{[\gamma, r] : \gamma \in \Gamma_{\text{ab}}, r \in \mathbb{R}\}$
 (b) *There exists $\lambda \in \mathbb{R}_{>0}$ and $f : \Gamma_{\text{ab}} \rightarrow \mathbb{R}$ such that*

$$\Gamma = \{[\gamma, \lambda(f(\gamma) + n)] : \gamma \in \Gamma_{\text{ab}}, n \in \mathbb{Z}\}.$$

Proof of Proposition 4. The full dimension condition guarantees that Γ_{ab} is a two dimensional closed subgroup of \mathbb{R}^2 , and the three possibilities given are all such closed subgroups. Likewise, the center of Γ is a non-trivial subgroup of \mathbb{R} , hence either \mathbb{R} or $\lambda \cdot \mathbb{Z}$ for some real $\lambda > 0$. It follows that the fiber over $\gamma \in \Gamma_{\text{ab}}$ is a translate of the center. Let v_1, v_2 be two linearly independent elements of the abelianization, and choose $g_1 = [v_1, z_1]$, $g_2 = [v_2, z_2]$ in Γ . The commutator $[g_1, g_2] = g_1 g_2 g_1^{-1} g_2^{-1}$ is bilinear in v_1, v_2 , is non-zero, and lies in the center. It follows that if one of v_1, v_2 may be scaled by a continuous parameter in the abelianization then the center is \mathbb{R} . \square

In the case in which Γ is discrete, a more detailed description of the fibers over the abelianization is available. In this case Γ_{ab} is a lattice in \mathbb{R}^2 . Let V be the covolume of this lattice

$$(7) \quad V = \text{vol}(\mathbb{R}^2 / \Gamma_{\text{ab}})$$

and let $\varepsilon : \Gamma_{\text{ab}} \rightarrow \{0, 1\}$ be the parity function defined by choosing a basis (v_1, v_2) for Γ_{ab} and setting

$$(8) \quad \forall n_1, n_2 \in \mathbb{Z}, \quad \varepsilon(n_1 v_1 + n_2 v_2) \equiv n_1 n_2 \pmod{2}.$$

It is straightforward to check that this definition is independent of the basis chosen.

Lemma 5. *Let*

$$A = \left\{ \underline{a} \in \mathbb{Z}^{|S|} : \sum_{g \in S} a_g g_{\text{ab}} = 0 \right\}.$$

The central fiber $\{x \in \mathbb{R} : [0, 0, x] \in \Gamma\}$ is, for some $L = \frac{V}{m} > 0$, $m \in \mathbb{Z}$

$$V \cdot \mathbb{Z} + \left\{ \sum_{g \in S} a_g \left(z_g - \frac{x_g y_g}{2} + \frac{\varepsilon_g V}{2} \right) : \underline{a} \in A \right\} = L \cdot \mathbb{Z}.$$

Proof. Recall (3) the multiplication rule for $\underline{w} \in S^N$ with $\overline{\underline{w}}^{(1)} = \overline{\underline{w}}^{(2)} = 0$,

$$\prod_{i=1}^N w_i = \left[\text{id}, \overline{\underline{w}}^{(3)} + H^*(\underline{w}) \right]; \quad \tilde{z} = z - \frac{1}{2}xy.$$

Choose a basis (v_1, v_2) for Γ_{ab} , and write each w_i as $w_i = s_i v_1 + t_i v_2$. Then $H^*(\underline{w})$ is some integral multiple of $\frac{V}{2}$, and the integral multiple is odd if and only if the number of w_i with odd parity is odd:

$$H^*(\underline{w}) = \frac{v_1 \wedge v_2}{2} \left(2 \sum_{i < j} s_i t_j + \sum_i s_i t_i \right).$$

□

Lemma 6. $\overline{\Gamma_0} = \Gamma$.

Proof. Write $\Gamma_{0,\text{ab}} = p(\Gamma_0)$ where p denotes projection to the abelianization G_{ab} . We first prove $\overline{\Gamma_{0,\text{ab}}} = \Gamma_{\text{ab}}$. Let $u \in \Gamma_{0,\text{ab}}$. We claim that for some $r < 0$, $ru \in \overline{\Gamma_{0,\text{ab}}}$. To see this, first choose $v \in \Gamma_{0,\text{ab}}$ such that $\langle u, v \rangle < 0$. If $v \neq ru$ then choose $w \in \Gamma_{0,\text{ab}}$ such that $\langle u, w \rangle < 0$ and $\langle \text{proj}_{u^\perp} v, \text{proj}_{u^\perp} w \rangle < 0$ (this is guaranteed since otherwise all of $\Gamma_{0,\text{ab}}$ is contained in a single half-plane determined by v). Form positive integer combinations of u, v, w to obtain $u' = ru$ with $r < 0$ in $\overline{\Gamma_{0,\text{ab}}}$. Note that this guarantees that $\overline{\Gamma_{0,\text{ab}}} \cap \{su : s \in \mathbb{R}\}$ is a group. As u was arbitrary $\overline{\Gamma_{0,\text{ab}}}$ is a group, hence equal to Γ_{ab} .

Let $0 < \epsilon < \frac{1}{4}$ be a fixed small parameter and choose x, x', y, y' in Γ_0 such that

$$p(x), p(x'), p(y), p(y') \approx e_1, -e_1, e_2, -e_2$$

where the approximation means to within distance ϵ . Take a word \underline{w} in $T = \{\text{id}, x, x', y, y'\}$ of length $4n$ with product approximating the identity in Γ_{ab} to within ϵ , which is such that each of x, x', y, y' appear $> (1 - O(\epsilon))n$ times in \underline{w} . The abelianization of the product is independent of the ordering of \underline{w} , but if the letters in \underline{w} appear in order y, x, y', x' then the central element is $< -(1 + O(\epsilon))n^2$, while if they appear in order y', x, y, x' then the central element is $> (1 + O(\epsilon))n^2$. Moving from an ordering of the first type to an ordering of the second by swapping generators one at a time changes the central element by $O(1)$ at each step. Let $\epsilon \downarrow 0$ to deduce that $\overline{\Gamma_0}$ contains positive and negative central elements, and hence that $\overline{\Gamma_0}$ is a group, equal to Γ . □

More quantitative structural statements are as follows.

Lemma 7. *Let μ be a measure on $\mathbb{H}(\mathbb{R})$, with abelianization μ_{ab} not supported on a lattice of \mathbb{R}^2 . If the Cramér condition holds for the measure μ_{ab} then it holds also for the measure on μ' on \mathbb{R} obtained by pushing forward $\mu_{\text{ab}} \otimes \mu_{\text{ab}}$ by $H^*(w_1, w_2)$.*

Proof. Let $\xi \in \mathbb{R}$, $|\xi| \geq 1$ and fix $w_2 \in \text{supp}(\mu_{\text{ab}})$, bounded away from 0. Write $H^*(w_1, w_2) = \frac{w_1 \wedge w_2}{2} = \frac{1}{2} w_1 \cdot \hat{w}_2$. The claim follows since $\left| \int e_{-\xi} (H^*(w_1, w_2)) d\mu_{\text{ab}}(w_1) \right|$ is bounded away from 1 uniformly in ξ and w_2 . □

Lemma 8. *Let μ be a measure on \mathbb{R}^2 of compact support, with support generating a subgroup of \mathbb{R}^2 of full dimension. If μ is lattice supported, assume that the co-volume of the lattice is at least 1. There is a constant $c = c(\mu) > 0$ such that, uniformly in $0 < \xi < \frac{1}{2}$, for $N = N(\xi) = \left\lfloor \frac{1}{\xi} \right\rfloor$,*

$$\left| \int_{\mathbb{R}^2 \times \mathbb{R}^2} e_{-\xi}(H^*(w_1, w_2)) d\mu^{*N}(w_1) d\mu^{*N}(w_2) \right| \leq 1 - c(\mu).$$

Proof. Standard application of the functional central limit theorem implies that $\frac{1}{N} H^*(w_1, w_2) d\mu^{*N}(w_1) d\mu^{*N}(w_2)$ converges to a non-zero density on \mathbb{R} as $N \rightarrow \infty$. \square

Normalize Haar measure on $\mathbb{H}(\mathbb{R})$ to be given in coordinates by $dg = dx dy dz$. The density of a Gaussian measure ν on $\mathbb{H}(\mathbb{R})$ can be understood as the rescaled limit of the density of a random walk with independent Gaussian inputs in the abelianization. Consider the distribution on the Heisenberg group given by $\nu_2 = [\eta(0, 1), \eta(0, 1), 0]$, which has projection to the abelianization given by a two dimensional symmetric standard normal distribution, and with trivial central fiber. The rescaled distribution $d_{\frac{1}{\sqrt{N}}} \nu_2^{*N}$ converges to a Gaussian measure ν_0 on $\mathbb{H}(\mathbb{R})$ as $N \rightarrow \infty$. Note that we have not included a covariance term, which can be accommodated with a linear change of coordinates. Also, we do not consider randomness in the central coordinate as it would scale only as \sqrt{N} , whereas the central coordinate has distribution at scale N .

Let $\alpha \in \mathbb{R}^2$ and $\xi \in \mathbb{R}$. Write the modified characteristic function of ν_0 as (recall $\tilde{z} = z - \frac{xy}{2}$)

$$I(\alpha, \xi) = \int_{g=[x,y,z] \in G} e_{-\alpha}(g_{\text{ab}}) e_{-\xi}(\tilde{z}) d\nu_0(g)$$

and

$$I(\alpha, \xi; N) = \int_{(\mathbb{R}^2)^N} e_{-\alpha} \left(\frac{\underline{x}}{\sqrt{N}} \right) e_{-\xi} \left(\frac{H^*(\underline{x})}{N} \right) d\nu_{2,\text{ab}}^{\otimes N}(\underline{x}).$$

In view of the multiplication rule (6), for $\|\alpha\|, |\xi| = O(1)$

$$\lim_{N \rightarrow \infty} I(\alpha, \xi; N) \rightarrow I(\alpha, \xi).$$

The following rate of convergence is given in Appendix A.

Theorem 9. *For all $\alpha \in \mathbb{R}^2$, $\xi \in \mathbb{R}$ such that $(1 + \|\alpha\|^2)(1 + \xi^2) < N$,*

$$I(\alpha, \xi; N) = \frac{1 + O\left(\frac{(1 + \|\alpha\|^2)(1 + \xi^2)}{N}\right)}{\exp\left(\frac{2\pi\|\alpha\|^2}{\xi \coth \pi \xi}\right) \cosh \pi \xi}.$$

In particular,

$$I(\alpha, \xi) = \frac{\exp\left(-\frac{2\pi\|\alpha\|^2}{\xi \coth \pi \xi}\right)}{\cosh \pi \xi}.$$

Remark. While $I(\alpha, \xi)$ characterizes the Gaussian measure, it does not behave well under convolution.

We make the following convention regarding rare events. Say that a sequence of measurable events $\{A_N\}_{N \geq 1}$ such that $A_N \subset S^N$ occurs *with high probability* (w.h.p.) if the complements satisfy the decay estimate,

$$\forall C \geq 0, \quad \mu^{\otimes N}(A_N^c) = O_C(N^{-C})$$

as $N \rightarrow \infty$. The sequence of complements is said to be *negligible*. A sequence of events $\{A_N\}$ which is negligible for $\mu^{\otimes N}$ is also negligible when $\mu^{\otimes N}$ is conditioned on a non-negligible sequence of events $\{B_N\}$.

Recall the classical local limit theorem for sums of independent random variables on \mathbb{R}^n , see e.g. [16].

Theorem 10 (Local limit theorem for \mathbb{R}^n). *Let μ be a lazy probability measure of mean zero, covariance matrix*

$$\underline{\sigma}^2 = \left(\int_{\mathbb{R}^n} x^{(i)} x^{(j)} d\mu(x) \right)_{i,j=1}^n,$$

compact support and such that $\Gamma = \overline{\langle \text{supp } \mu \rangle}$ satisfies \mathbb{R}^n/Γ is compact. Let \mathcal{F} be a fundamental domain for \mathbb{R}^n/Γ , and let $\chi_{\mathcal{F}} = \frac{1_{\mathcal{F}}}{\text{vol}(\mathcal{F})}$, where $\chi_{\mathcal{F}} = \delta_0$ in the case that $\Gamma = \mathbb{R}^n$. Denote by $\eta(0, \underline{\sigma})$ the standard normal distribution with covariance matrix $\underline{\sigma}^2$. Let $f \in C_c^\infty(\mathbb{R}^n)$, $f \geq 0$ and, given $x \in \mathbb{R}^n$, write $T_x f(y) = f(y - x)$ for the translated function. For $N \geq 1$, uniformly in $x \in \mathbb{R}^n$, for all $0 < \epsilon < \frac{1}{3}$,

$$\begin{aligned} \langle T_x f * \chi_{\mathcal{F}}, \mu^{*N} \rangle &= (1 + o(1)) \langle T_x f, \eta(0, \sqrt{N} \underline{\sigma}) \rangle \\ &\quad + O_\epsilon \left(\exp \left(-N^{\frac{1}{3}-\epsilon} \right) \right). \end{aligned}$$

If the Cramér condition is satisfied,

$$\sup_{\substack{\lambda \in \widehat{\mathbb{R}^n_{\text{nl}}} \\ |\lambda| > 1}} \left| \int_{\mathbb{R}^n} e_{-\lambda}(x) d\mu(x) \right| < 1$$

then, for all $N \geq 1$, uniformly in $x \in \mathbb{R}^n$, for all $f \in C_c^\infty(\mathbb{R}^n)$, $f \geq 0$ and for all $0 < \epsilon < \frac{1}{3}$,

$$\begin{aligned} \langle T_x f * \chi_{\mathcal{F}}, \mu^{*N} \rangle &= \left(1 + O \left(N^{-\frac{1}{2}} + \frac{|x|^3}{N^2} \right) \right) \langle T_x f, \eta(0, \sqrt{N} \underline{\sigma}) \rangle \\ &\quad + O_\epsilon \left(\exp \left(-N^{\frac{1}{3}-\epsilon} \right) \right). \end{aligned}$$

Proof sketch. Suppose first that $\overline{\langle \text{supp } \mu \rangle} = \mathbb{R}^n$. It may be assumed that f has been smoothed at scale $\exp \left(-N^{\frac{1}{3}} \right)$ and that $|x| \ll N^{\frac{2}{3}}$.

First assume the Cramér condition. Write

$$\langle T_x f, \mu^{*N} \rangle = \int_{\mathbb{R}^n} \hat{f}(\xi) e_{-\xi}(x) \overline{\hat{\mu}(\xi)}^N d\xi.$$

The integral is split into the ranges $|\xi| \leq N^{-\frac{1}{3}-\varepsilon}$, $N^{-\frac{1}{3}-\varepsilon} < |\xi| \leq \varepsilon$ and $\varepsilon < |\xi|$.

On the first range write

$$\forall \xi \in \mathbb{C}, |\xi| \ll N^{-\frac{1}{3}}, \quad \hat{\mu}(\xi)^N = \exp(-2\pi^2 N \xi^t \sigma^2 \xi) (1 + O(N|\xi|^3)),$$

Make the linear change of coordinates

$$\xi := \sqrt{N} \sigma \xi - \frac{i \sigma^{-1} x}{\sqrt{N}},$$

and shift the integral to be on the real plane. Bounding

$$\sup_{\xi_0 \in \mathbb{R}^n} \left| \hat{f} \left(\xi_0 + i \frac{\sigma^{-2} x}{N} \right) \right| \ll \|f\|_1$$

gives

$$\begin{aligned} & \left\langle T_x f, \eta \left(0, \sqrt{N} \underline{\sigma} \right) \right\rangle + O_\epsilon \left(\|f\|_1 \exp \left(-N^{\frac{1}{3}-\epsilon} \right) \right) \\ & + O \left(\frac{\|f\|_1}{\sqrt{\det \sigma^2} N^{\frac{n}{2}}} \exp \left(\frac{-x^t (\sigma^2)^{-1} x}{2} \right) \left(\frac{1}{\sqrt{N}} + \frac{|x|^3}{N^2} \right) \right). \end{aligned}$$

In the intermediate range $N^{-\frac{1}{3}-\varepsilon} < |\xi| \leq \varepsilon$, $\xi \in \mathbb{R}$, bound simply

$$1 - |\hat{\mu}(\xi)| \gg \xi^t \sigma^2 \xi,$$

which gives another error bounded by

$$O_\epsilon \left(\|f\|_1 \exp \left(-N^{\frac{1}{3}-\epsilon} \right) \right).$$

On the remaining range invoke Cramér's condition to obtain a bound, for some $\epsilon > 0$,

$$O_\epsilon \left(\exp(-\epsilon N) \right),$$

which is smaller than the bound claimed.

In the case where Cramér is not assumed, approximate f in L^1 from above and below by band-limited functions f_+ and f_- and repeat the argument.

In the case where $\overline{\langle \text{supp } \mu \rangle}$ has a lattice component of dimension d , d dimensions of the integral are replaced by integrals over a torus, and compactness of the torus replaces the Cramér condition on this component.

□

3. PROOF OF THEOREM 1

The argument differs slightly according as the abelianized walk is lattice or non-lattice. We initially discuss only the case in which the walk takes place within a discrete subgroup of $\mathbb{H}(\mathbb{R})$, and then describe the necessary modifications needed to handle the other cases.

Let v_1, v_2 be a basis for Γ_{ab} and choose representatives $[v_1, z_1], [v_2, z_2] \in \Gamma$. Recall (7), $V = |v_1 \wedge v_2|$ and the notation from Lemma 5, for $[x, y, z] \in G$, $\tilde{z} = z - \frac{1}{2}xy$. Given an element $v = n_1v_1 + n_2v_2$ of Γ_{ab} , the fiber above v in Γ is, as a collection of third coordinates, the coset

$$n_1\tilde{z}_1 + n_2\tilde{z}_2 + \frac{n_1n_2}{2}v_1^{(1)}v_2^{(2)} + \frac{V}{m} \cdot \mathbb{Z}.$$

Given $\underline{n} = [n_1, n_2, n_3] \in \mathbb{Z}^3$, set

$$g_{\underline{n}} = \left[n_1v_1 + n_2v_2, n_1\tilde{z}_1 + n_2\tilde{z}_2 + \frac{n_1n_2}{2}v_1^{(1)}v_2^{(2)} + n_3\frac{V}{m} \right].$$

In order to prove Theorem 1 in the lattice case, it thus suffices to prove the following estimate.

Proposition 11. *For each $\underline{n} = (n_1, n_2, n_3) \in \mathbb{Z}^3$,*

$$(9) \quad p(n_1, n_2, n_3) := \mu^{*N}(\{g_{\underline{n}}\}) = \frac{V^2}{mN^2} \cdot \frac{d\nu}{dg} \left(d_{\frac{1}{\sqrt{N}}} g_{\underline{n}} \right) + O\left(N^{-\frac{5}{2}}\right).$$

Introduce two collections of words: $W_N = S^N$ and

$$W_N^{\underline{n}} = \left\{ \underline{w} \in W_N : \prod_{i=1}^N w_i \in [0, 0, V \cdot \mathbb{Z}] \cdot g_{\underline{n}} \right\}.$$

Recall the parity function ε from (8) and define $\overline{w}^{(\varepsilon)} = \sum_{i=1}^N \varepsilon(\underline{w}_{i,\text{ab}})$. The condition defining $W_N^{\underline{n}}$ is a condition on the pair $(\underline{w}, \overline{w}^{(\varepsilon)})$ as a vector in a lattice walk in $\mathbb{R}^3 \times \mathbb{Z}/2\mathbb{Z}$; the first two coordinates are fixed, and, within the fiber over these coordinates, the last two variables are fixed in an index m coset of the lattice. In particular, the lattice case of the local limit theorem on \mathbb{R}^4 gives that (write $n_1v_1 + n_2v_2 = v$)

$$\begin{aligned} \mu^{\otimes N}(W_N^{\underline{n}}) &= \frac{V(1+o(1))}{2\pi mN\sqrt{\det \underline{\sigma}^2}} \exp\left(-\frac{v^t(\underline{\sigma}^2)^{-1}v}{2N}\right) \\ &\quad + O_\epsilon\left(\exp\left(-N^{\frac{1}{3}-\epsilon}\right)\right), \end{aligned}$$

where $\underline{\sigma}^2$ denotes the covariance matrix of the walk projected orthogonally onto the first two coordinates. In particular, it may be assumed that

$$\|v\| \ll \sqrt{N \log N},$$

a condition which guarantees that $W_N^{\underline{n}}$ is non-negligible.

It suffices to determine the conditional probability $\mathbf{Prob}(\{g_{\underline{n}}\}|W_N^{\underline{n}})$. In what follows, $\mu^{\otimes N}$ is abbreviated \mathbb{U}_N and the conditional measure $\mu^{\otimes N}(\cdot|W_N^{\underline{n}})$ is abbreviated $\mathbb{U}_N^{\underline{n}}(\cdot)$. Recalling the multiplication rule

$$\prod_{i=1}^N [w_i^{(1)}, w_i^{(2)}, w_i^{(3)}] = \left[\underline{w}^{(1)}, \underline{w}^{(2)}, \underline{w}^{(3)} + \frac{1}{2} \underline{w}^{(1)} \underline{w}^{(2)} + H^*(\underline{w}) \right]$$

set $z_{\underline{n}} = n_1 \tilde{z}_1 + n_2 \tilde{z}_2 + n_3 \frac{V}{m}$. Also, write

$$\rho(t) = \chi_{[-\frac{1}{2}, \frac{1}{2}]}(t), \quad \rho_V(t) = \frac{1}{V} \rho\left(\frac{t}{V}\right).$$

Shifting the left and right integrand by $\frac{1}{2} \underline{w}^{(1)} \underline{w}^{(2)}$,

$$(10) \quad \mathbb{U}_N^{\underline{n}}(\{g_{\underline{n}}\}) = \left\langle V \rho_V * \delta_{z_{\underline{n}}}, \rho_V * \mathbf{E}_{\mathbb{U}_N^{\underline{n}}} \left[\delta_{\underline{w}^{(3)} + H^*(\underline{w})} \right] \right\rangle,$$

where

$$\langle f, g \rangle = \int_{\mathbb{R}} f g.$$

3.1. Reduction to central limit theorem. In frequency space,

$$(11) \quad \mathbb{U}_N^{\underline{n}}(\{g_{\underline{n}}\}) = V \int_{\mathbb{R}/\frac{1}{V}\mathbb{Z}} e_{-\xi}(z_{\underline{n}}) \mathbf{E}_{\mathbb{U}_N^{\underline{n}}} \left[e_{\xi} \left(H^*(\underline{w}) + \underline{w}^{(3)} \right) \right] d\xi.$$

The following two lemmas reduce to a quantitative central limit theorem by truncating frequency space to the scale of the distribution.

Lemma 12. *For any $A > 0$ there is $C = C(A) > 0$ such that if $\|V\xi\|_{\mathbb{R}/\mathbb{Z}} \geq \frac{C \log N}{N}$,*

$$\left| \mathbf{E}_{\mathbb{U}_N^{\underline{n}}} \left[e_{\xi} \left(H^*(\underline{w}) + \underline{w}^{(3)} \right) \right] \right| \leq N^{-A}.$$

Proof. Choose $k = k(\xi)$ according to the rule

$$k(\xi) = 1, \quad |\xi| > \frac{1}{10V}, \quad \left\lfloor \frac{1}{2V|\xi|} \right\rfloor, \quad |\xi| \leq \frac{1}{10}.$$

Let $N' = \lfloor \frac{N}{2k} \rfloor$. The group $G_k = C_2^{N'}$ acts on strings of length N with j th factor exchanging the order of the substrings of length k ending at $(2j-1)k$ and $2jk$. The action preserves $W_N^{\underline{n}}$, since only the value of $H^*(\underline{w})$ is altered, and it changes by a multiple of V . Given string \underline{w} , write $\hat{\underline{w}}$ for the string of length $2N'$ with j th entry given by

$$\hat{w}_j = \sum_{i=1}^k w_{(j-1)k+i}.$$

Write

$$H^*(\underline{w}) = H_k^1(\underline{w}) + H_k^2(\underline{w}), \quad H_k^2(\underline{w}) = \sum_{j=1}^{N'} H^*(\hat{w}_{2j-1}, \hat{w}_{2j}).$$

H_k^1 is invariant under G_k . One has

$$\begin{aligned} & \mathbf{E}_{\mathbb{U}_N^n} \left[e_\xi \left(H^*(\underline{w}) + \overline{\underline{w}}^{(3)} \right) \right] \\ &= \mathbf{E}_{\mathbb{U}_N^n} \left[\mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H^*(\mathcal{T} \cdot \underline{w}) + \overline{\underline{w}}^{(3)} \right) \right] \right] \\ &= \mathbf{E}_{\mathbb{U}_N^n} \left[e_\xi \left(H_k^1(\underline{w}) + \overline{\underline{w}}^{(3)} \right) \mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H_k^2(\mathcal{T} \cdot \underline{w}) \right) \right] \right], \end{aligned}$$

and, therefore,

$$\begin{aligned} \left| \mathbf{E}_{\mathbb{U}_N^n} \left[e_\xi \left(H^*(\underline{w}) + \overline{\underline{w}}^{(3)} \right) \right] \right| &\leq \mathbf{E}_{\mathbb{U}_N^n} \left[\left| \mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H_k^2(\mathcal{T} \cdot \underline{w}) \right) \right] \right| \right] \\ &\leq N^{O(1)} \mathbf{E}_{\mathbb{U}_N} \left[\left| \mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H_k^2(\mathcal{T} \cdot \underline{w}) \right) \right] \right| \right]. \end{aligned}$$

One checks

$$\mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H_k^2(\mathcal{T} \cdot \underline{w}) \right) \right] = \prod_{j=1}^{N'} \cos(2\pi \xi H^*(\hat{w}_{2j-1}, \hat{w}_{2j})).$$

Let $\delta = \delta(\mu) > 0$ be a small parameter. Let $E_k(j)$ be the event $\|\xi \cdot H^*(\hat{w}_{2j-1}, \hat{w}_{2j})\|_{\mathbb{R}/\mathbb{Z}} \geq \delta$. Choosing \underline{w} according to \mathbb{U}_N , for $j < N'$ the events $E_k(j)$ are i.i.d. Furthermore, if δ is sufficiently small then Lemma 8 implies that $E_k(j)$ occurs with positive probability uniformly in k . Let $C > 0$ be a small constant, and let E_{bad} be the event $\sum_j E_k(j) < CN'$. If C is sufficiently small then

$$\mathbb{U}_N(E_{\text{bad}}) \leq \exp(-CN')$$

while on E_{bad}^c there is $C' > 0$ such that

$$\mathbf{E}_{\mathcal{T} \in G_k} \left[e_\xi \left(H_k^2(\mathcal{T} \cdot \underline{w}) \right) \right] \leq \exp(-C'N'),$$

completing the estimate. \square

The above lemma permits truncation of the integral in ξ at $|\xi| \ll \frac{\log N}{N}$. Next the conditioning is removed by fixing the abelianized variables in frequency space.

Recall $mL = V$. The conditions for $\underline{w} \in W_N^n$ are, for some $x \bmod m$,

$$\overline{\underline{w}}_{\text{ab}} = n_1 v_1 + n_2 v_2, \quad \overline{\underline{w}}^{(3)} + \frac{mL}{2} \overline{\underline{w}}^{(\varepsilon)} \in xL + mL\mathbb{Z}.$$

This may be imposed as

$$\begin{aligned} (12) \quad \mathbf{1}(\underline{w} \in W_N^n) &= \frac{1}{m} \sum_{a \bmod m} e_{a,m} \left(-x + \frac{1}{L} \left(\overline{\underline{w}}^{(3)} + \frac{mL}{2} \overline{\underline{w}}^{(\varepsilon)} \right) \right) \\ &\quad \times \int_{(\mathbb{R}/\mathbb{Z})^2} e_\alpha \left(-[n_1, n_2]^t + [v_1, v_2]^{-1} \overline{\underline{w}}_{\text{ab}} \right) d\alpha. \end{aligned}$$

Let $(\alpha')^t = \alpha^t [v_1, v_2]^{-1}$.

Lemma 13. *Let $A, \epsilon > 0$ and $0 \leq \|V\xi\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{C \log N}{N}$ where C is as in Lemma 12. For all N sufficiently large, if either $a \not\equiv 0 \pmod{m}$ or $\|\alpha\|_{\mathbb{R}^2/\mathbb{Z}^2} \geq N^{\epsilon-\frac{1}{2}}$, then*

$$\mathbf{E}_{\mathbb{U}_N} \left[e_{a,m} \left(\frac{\overline{w}^{(3)}}{Lm} + \frac{1}{2} \overline{w}^{(\epsilon)} \right) e_{\alpha'}(\overline{w}_{\text{ab}}) e_{\xi} \left(H^*(\underline{w}) + \overline{w}^{(3)} \right) \right] \leq N^{-A}.$$

Proof. Let $\varepsilon = \frac{\epsilon}{2}$ and let $N' = \lfloor N^{1-\varepsilon} \rfloor$. Let \underline{w}_0 be \underline{w} truncated at N' and let \underline{w}_t be the remainder of \underline{w} so that \underline{w} is the concatenation $\underline{w}_0 \oplus \underline{w}_t$. Treat \underline{w}_t as fixed and assume that $\|\underline{w}_t\| \leq \sqrt{N} \log N$, which holds w.h.p. Write

$$H^*(\underline{w}) = H^*(\underline{w}_0) + H^*(\overline{w}_0, \overline{w}_t) + H^*(\underline{w}_t).$$

Let $E_k(x)$ denote the degree k Taylor expansion of $e_1(x)$ about 0. Let $k = k(A, \epsilon)$ be sufficiently large so that $E_k(\xi H^*(\underline{w}_0)) - e_{\xi}(H^*(\underline{w}_0)) \leq \frac{1}{2N^A}$ holds w.h.p. The error in making this approximation may be bounded by taking a moment higher than k of $\xi H^*(\underline{w}_0)$, and is negligible. It thus suffices to estimate instead

$$\begin{aligned} \mathbf{E}_{\mathbb{U}_{N'}} \left[e_{a,m} \left(\frac{\overline{w}_0^{(3)}}{mL} + \frac{m}{2} \overline{w}_0^{(\epsilon)} \right) e_{\alpha'}(\overline{w}_{0,\text{ab}}) e_{\xi} \left(H^*(\overline{w}_0, \overline{w}_t) + \overline{w}_0^{(3)} \right) \right. \\ \left. \times E_k(\xi H^*(\underline{w}_0)) \right]. \end{aligned}$$

Expand E_k into $\text{Poly}(N)$ terms, each depending on boundedly many indices from \underline{w}_0 . Expectation over the remaining terms factors as a product which is exponentially small in a power of N , hence negligible. \square

3.2. Quantitative Gaussian approximation. In the range $\|\alpha\| \leq N^{\epsilon-\frac{1}{2}}$, $|\xi| \ll \frac{\log N}{N}$, expectation with respect to μ is replaced with expectation taken over a measure with projection to the abelianization given by a Gaussian of the same covariance matrix as μ_{ab} . The modified characteristic function in the Gaussian case is evaluated precisely in Theorem 9, which finishes the proof.

Lemma 14. *Write σ^2 for the covariance matrix of μ_{ab} . Let*

$$\sigma \alpha' = \alpha'', \quad \delta = \det \sigma.$$

There exists $c > 0$ such that, for $\|\alpha\| \leq N^{\epsilon-\frac{1}{2}}$ and $|\xi| \leq \frac{C \log N}{N}$,

$$\begin{aligned}
& \mathbf{E}_{\mathbb{U}_N} \left[e_{\alpha'}(\underline{w}_{\text{ab}}) e_{\xi} \left(H^*(\underline{w}) + \overline{\underline{w}}^{(3)} \right) \right] \\
&= e_{\xi} \left(N \overline{\underline{z}} \right) I \left(N^{\frac{1}{2}} \alpha'', N \delta \xi; N \right) + O(N^{\epsilon-1}) \\
(13) \quad &+ \min \left[O \left(\|\alpha\| (1 + N \|\alpha\|^2) (1 + N^3 |\xi|^3) \right) \right. \\
&\quad \times \exp \left(-c \left(\|\alpha\|^2 \min \left(N, \frac{1}{N |\xi|^2} \right) \right) \right), \\
(14) \quad &\left. O \left(N \|\alpha\|^3 + N^{\frac{1}{2}} |\xi| (1 + N^2 |\xi|^2) \right) \exp(-c N |\xi|) \right].
\end{aligned}$$

Proof. Without loss, let $\overline{\underline{z}} = 0$. Then $e_{\xi}(\overline{\underline{w}}^{(3)})$ may be removed by Taylor expansion at the end of the argument, the details are omitted.

Let ν be a centered Gaussian on \mathbb{R}^2 with covariance matrix equal to that of μ_{ab} . Since the expectation no longer depends upon the third coordinate, write $\mu_{\text{ab}}^{\otimes N}$ in place of \mathbb{U}_N . For $0 \leq n \leq N$ consider the measure $\mu_n = \mu_{\text{ab}}^{\otimes n} \otimes \nu^{\otimes N-n}$ in which the first n coordinates are i.i.d. with measure μ_{ab} and last $N - n$ coordinates are i.i.d. ν . Write

$$E_n = \mathbf{E}_{\mu_n} [e_{\alpha'}(\underline{w}_{\text{ab}}) e_{\xi}(H^*(\underline{w}))].$$

Since $E_0 = I \left(N^{\frac{1}{2}} \alpha'', N \delta \xi; N \right)$ (the expectation is real) it suffices to bound the difference

$$E_N - E_0 = \sum_{j=1}^N E_j - E_{j-1}.$$

To bound $E_j - E_{j-1}$, write E_j as an N -fold integral, and move integration with respect to the j th coordinate inside. This inner expectation is the complex conjugate of the characteristic function of μ_{ab} at frequency

$$\alpha_j(\underline{w}) = \alpha' + \frac{\xi}{2} \left[\sum_{i \neq j} (-1)^{\delta(i < j)} w_i^{(2)}, \sum_{i \neq j} (-1)^{\delta(i > j)} w_i^{(1)} \right]^t.$$

By Taylor expansion,

$$(15) \quad \mathbf{E}_{\mu_{\text{ab}}} [e_{\alpha_j(\underline{w})}(w_j)] = \mathbf{E}_{\nu} [e_{\alpha_j(\underline{w})}(w_j)] (1 + T(\alpha_j(\underline{w})) + O(\|\alpha_j(\underline{w})\|^4))$$

where $T(\cdot)$ is a degree 3 polynomial. Since

$$\mathbf{E}_{\mu_j} [\|\alpha_j(\underline{w})\|^4] \ll N^{\epsilon-2}$$

it suffices to bound the error that results from T .

The error from T_j is bounded in one of two ways depending upon the relative sizes of $\|\alpha\|$ and $|\xi|$. First, by an argument analogous to the argument in the case of Lemma 12 in which the order of blocks

of length $k \sim \frac{1}{|\xi|}$ are swapped in a fixed string, we obtain an error, summed over all j , of $O\left(\exp(-cN|\xi|)(N^{\frac{5}{2}}|\xi|^3 + N\|\alpha\|^3)\right)$, which is the bound claimed in (14). To make this argument, form G'_k by omitting from G_k the factor which moves index j . Then T_j is invariant under G'_k . Averaged with respect to \mathbb{U}_N , the expected size of $|T_j|$ can be separated from the expected size of $E_{\tau \in G'_k}[e_\xi(H^*(\tau \cdot \underline{w}))]$ by Cauchy-Schwarz.

To obtain decay in $\|\alpha\|$, iterate the estimate (15) to find

$$\begin{aligned} & E_N + O(N^{\epsilon-1}) \\ &= \mathbf{E}_{\nu^{\otimes N}} \left[e_{\alpha'}(\underline{w}_{\text{ab}}) e_\xi(H^*(\underline{w})) \left(1 + \sum_j T(\alpha_j(\underline{w})) \right) \right]. \end{aligned}$$

The degree 3 polynomial $T = \sum_j T_j(\alpha_j(\underline{w}))$ consists in monic monomials of which

- i. $O(N)$ are constant in \underline{w} and cubic in α
- ii. $O(N^2)$ are linear in $\xi \underline{w}$ and quadratic in α
- iii. $O(N^3)$ are quadratic in $\xi \underline{w}$ and linear in α . Of these $O(N^2)$ have a repeated factor from \underline{w}
- iv. $O(N^4)$ are cubic in $\xi \underline{w}$. Of these, $O(N^3)$ have a repeated factor from \underline{w} .

Given a typical monomial M of T , write $\omega(M)$ for the number of variables from \underline{w} which are odd degree in M . Consider expectation

$$E_M = \mathbf{E}_{\nu^{\otimes N}} [M e_{\alpha'}(\underline{w}_{\text{ab}}) e_\xi(H^*(\underline{w}))].$$

Let $C > 0$ be a small constant, and let $N' = \min\left(N, \frac{C}{|\xi|}\right)$. Let \underline{w}_0 be the initial string of \underline{w} of length N' and assume that this includes any variables from M ; the general case may be handled by a trivial modification. Write $\underline{w} = \underline{w}_0 \oplus \underline{w}_t$ so that \underline{w}_t contains the remaining variables. Treat \underline{w}_t as fixed and average over \underline{w}_0 . After performing a rotation of \mathbb{R}^2 simultaneously in each coordinate one may assume that α' is proportional to $[1, 1]^t$. Write

$$H^*(\underline{w}) = H^*(\underline{w}_0) + H^*(\underline{w}_0, \underline{w}_t) + H^*(\underline{w}_t).$$

Write $\hat{\alpha} = \alpha' + \frac{\xi}{2} \left[\underline{w}_t^{(2)}, -\underline{w}_t^{(1)} \right]^t$. Expand $e_\xi(H^*(\underline{w}_0))$ in Taylor series to degree $L := \lfloor N^{2\epsilon} \rfloor$. The error in doing so is bounded by taking the next even moment, and is negligible, see the argument below.

It remains to bound

$$\begin{aligned} & \sum_{\ell=0}^L \frac{(2\pi|\xi|)^\ell}{\ell!} |\mathbf{E}_{\nu^{\otimes N'}} [M e_{\hat{\alpha}}(\underline{w}_0) H^*(\underline{w}_0)^\ell]| \\ & \leq \sum_{l=0}^L \frac{(2\pi|\xi|)^\ell}{\ell!} \sum_{\underline{m}, \underline{n} \in [N']^\ell} |E_{\nu^{\otimes N'}} [M e_{\hat{\alpha}}(\underline{w}_0) w_{m_1}^{(1)} \cdots w_{m_\ell}^{(1)} w_{n_1}^{(2)} \cdots w_{n_\ell}^{(2)}]|. \end{aligned}$$

Perform expectation over all variables not appearing among $\underline{m}, \underline{n}, M$. This obtains a factor of, for some $c > 0$, $\exp(-c\|\hat{\alpha}\|^2 N')$. In the remaining variables, Taylor expand the exponential $e_{\hat{\alpha}}$. The dominant term then comes from the least even term in each variable. Summing over $\underline{m}, \underline{n}$, the dominant contribution comes with each m_i, n_j appearing once and not overlapping M . This obtains a bound of

$$\begin{aligned} & \ll \exp(-c\|\hat{\alpha}\|^2 N') \|\hat{\alpha}\|^{\omega(M)} \sum_{\ell=0}^L \frac{(O(1)\|\hat{\alpha}\|^2 |\xi| (N')^2)^\ell}{\ell!} \\ & \ll \exp(-\|\hat{\alpha}\|^2 (-cN' + O(|\xi|(N')^2))) \|\hat{\alpha}\|^{\omega(M)}. \end{aligned}$$

If the constant C is chosen sufficiently small, this is bounded by, for some $c > 0$

$$\exp\left(-c\|\hat{\alpha}\|^2 \min\left(N, \frac{C}{|\xi|}\right)\right) \|\hat{\alpha}\|^{\omega(M)}.$$

The resulting bound is acceptable unless $\|\hat{\alpha}\| < c\|\alpha\|$ for a small constant c . In this case one obtains $\|\xi \underline{w}_t\| \gg \|\alpha\|$ which is an event which occurs with probability $\ll \exp\left(-c\frac{\|\alpha\|^2}{N\xi^2}\right)$, which is again satisfactory.

The bound claimed in (13) follows on considering the description of monomials M given above. \square

Proof of Theorem 1, lattice case. Combining Lemmas 12, 13, and 14 obtains, for any $A > 0$, $0 < \epsilon < \frac{1}{4}$, for some $c > 0$

$$\begin{aligned} & p(n_1, n_2, n_3) + O_A(N^{-A}) \\ &= \frac{V}{m} \iiint_{\substack{\|\alpha\| \leq N^{\epsilon - \frac{1}{2}} \\ |\xi| \leq \frac{\log N}{N}}} e_{-\alpha}([n_1, n_2]^t) e_{-\xi}(n_1 \tilde{z}_1 + n_2 \tilde{z}_2 + n_3 L - N \tilde{z}) \\ & \quad \times \left[I(\sqrt{N}\alpha'', N\delta\xi; N) + O(E) \right] d\alpha d\xi, \end{aligned}$$

where the error term E satisfies the estimate of Lemma 14. Over the range of integration this integrates to $O(N^{-\frac{5}{2}})$.

Again on the range of integration,

$$I(\sqrt{N}\alpha'', N\delta\xi; N) = I(\sqrt{N}\alpha'', N\delta\xi) (1 + O(N^{\epsilon-1})),$$

see Theorem 9. Making a change of variables and extending the integral to \mathbb{R}^3 obtains

$$\begin{aligned} & p(n_1, n_2, n_3) + O\left(N^{-\frac{5}{2}}\right) \\ &= \frac{V^2}{m\delta^2 N^2} \int_{\mathbb{R}^3} e_{-\alpha} \left((\sigma^2)^{-\frac{1}{2}} \left[\frac{n_1 v_1 + n_2 v_2}{\sqrt{N}} \right] \right) \\ & \quad \times e_{-\xi} \left(\frac{1}{\delta} \left(\frac{n_1 \tilde{z}_1 + n_2 \tilde{z}_2 + n_3 L}{N} - \tilde{z} \right) \right) I(\alpha, \xi) d\alpha d\xi. \end{aligned}$$

The right hand side is the Gaussian density of the limit theorem.

□

Proof of Corollary 2. The required probability is $p(0, 0, 0)$. One has $m = V = 1$ and $\delta^2 = \frac{4}{25}$. The value $\frac{25}{16}$ is obtained since

$$\int_{\mathbb{R}^3} I(\alpha, \xi) d\alpha d\xi = \frac{1}{4}.$$

□

4. NON-LATTICE MEASURES

The most straightforward modification is to the case that the abelianized measure is discrete but the central fiber is not. In this case the point g_n is replaced with a point $g_{n_1, n_2, z} = [n_1 v_1 + n_2 v_2, z]$ where z is a real parameter. The function $V\rho_V$ of (10) is replaced with a smooth function f of compact support. The integral over ξ in (11) becomes an integral over \mathbb{R} instead of a torus. This integral is split into the product of an integral and a sum by setting $\xi = \xi_0 + \frac{k}{V}$ with $|\xi_0| \leq \frac{1}{2V}$ and $k \in \mathbb{Z}$. Lemma 12 handles the case that $|\xi_0|$ is large. When Cramér is assumed, $k \neq 0$ is analogous to $a \neq 0$ in Lemma 13. If Cramér is not assumed then f is approximated above and below in L^1 with band-limited functions so that only finitely many k need be considered, but in this case a rate is not obtained. The remainder of the argument goes through as before.

In the case where the abelianized walk has a continuous factor the details are only slightly more involved. In this case the fibered distribution is also dense in \mathbb{R} . When both abelianized coordinates are continuous test against functions of type

$$f([x, y, z]) = F\left(x - x_0, y - y_0, z - \frac{xy}{2} - Ax - By - z_0\right)$$

where F is a smooth non-negative function of compact support and x_0, y_0, z_0, A, B are real parameters. For fixed \underline{w} , the first two coordinates in F are fixed and the corresponding function of the third coordinate replaces ρ_V in the argument above. This is essentially the only change to (10). When Cramér is assumed for μ_{ab} it holds also for $H^*(w_1, w_2) d\mu_{\text{ab}}(w_1) d\mu_{\text{ab}}(w_2)$ as a distribution on \mathbb{R} , and this is sufficient to apply Lemma 12 to truncate to small ξ . Open F via Fourier transform, again with α and β dual to the first two coordinates. If the Cramér condition holds it gives a rate in truncating α and β . Note that the shifts A, B introduce linear phases in the α and β integrals but this does not alter the estimation of the error, in which the integrand is bounded in absolute value. In the case where the Cramér condition is not assumed, approximate F from above and below to within $\epsilon > 0$ in L^1 with finite sums

$$\sum_{j=1}^J \rho_{1,j} \otimes \rho_{2,j} \otimes \rho_{3,j}$$

where each $\rho_{i,j}$ has Fourier transform supported in an interval of length $O_\epsilon(1)$. The proof goes through as before, but does not give an effective rate.

The case in which one of the two abelianized variables is discrete is handled by replacing one real Fourier integral with an integration on a torus.

5. RANDOM WALK ON $N_n(\mathbb{Z})$, PROOF OF THEOREM 3

The case $n = 2$ is classical and the case $n = 3$ may be deduced from Theorem 1, so we consider $n \geq 4$.

Let $M : \mathbb{Z}^{n-1} \rightarrow N_n(\mathbb{Z})$ be the map

$$M : \mathbb{Z}^{n-1} \ni v = \begin{pmatrix} v^{(1)} \\ v^{(2)} \\ \vdots \\ v^{(n-1)} \end{pmatrix} \mapsto \begin{pmatrix} 1 & v^{(1)} & 0 & \cdots & 0 \\ 0 & 1 & v^{(2)} & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ & 0 & 0 & 1 & v^{(n-1)} \\ 0 & \cdots & & 0 & 1 \end{pmatrix}$$

Recall that, given $m \in N_n$, $Z(m)$ is the central coordinate (upper right corner entry). Given sequence of vectors $\underline{v} = \{v_i\}_{i=1}^N \in (\mathbb{Z}^{n-1})^N$ the central coordinate satisfies the product rule

$$Z\left(\prod_{i=1}^N M(v_i)\right) = \sum_{1 \leq i_1 < i_2 < \dots < i_{n-1} \leq N} v_{i_1}^{(1)} v_{i_2}^{(2)} \cdots v_{i_{n-1}}^{(n-1)}.$$

Write

$$Z_n^N = \sum_{1 \leq i_1 < i_2 < \dots < i_{n-1} \leq N} e_{i_1}^{(1)} \otimes \cdots \otimes e_{i_{n-1}}^{(n-1)}$$

for the corresponding tensor. $Z_{n,\mu}^N$ denotes the measure on \mathbb{Z} obtained by pushing forward measure μ on \mathbb{Z}^{n-1} via M to measure $\tilde{\mu}$ on $N_n(\mathbb{Z})$, then obtaining $\langle Z, \tilde{\mu}^{*N} \rangle$. Equivalently, $Z_{n,\mu}^N$ is the distribution of Z_n^N evaluated on N vectors v_i drawn i.i.d. from μ .

Given a probability measure ν on \mathbb{Z} and prime p , Cauchy-Schwarz and Plancherel give

$$(16) \quad \sum_{x \bmod p} \left| \nu(x) - \frac{1}{p} \right| \leq \left(\sum_{0 \neq \xi \bmod p} \left| \hat{\nu}\left(\frac{\xi}{p}\right) \right|^2 \right)^{\frac{1}{2}}$$

where

$$\hat{\nu}(\alpha) = \sum_{n \in \mathbb{Z}} e_{-\alpha}(n) \nu(n).$$

Theorem 3 thus reduces to the following estimate on the characteristic function of Z_n^N .

Proposition 15. *Let $n \geq 4$ and let μ be a measure on \mathbb{Z}^{n-1} satisfying the same conditions as in Theorem 3. There exists constant $C > 0$ such that for all $N > 0$ and all $0 < |\xi| \leq \frac{1}{2}$*

$$\left| \hat{Z}_{n,\mu}^N(\xi) \right| \ll \exp \left(-CN |\xi|^{\frac{2}{n-1}} \right).$$

Deduction of Theorem 3. Recall $N = cp^{\frac{2}{n-1}}$ and let $c \geq 1$. Apply the upper bound of Proposition 15. By (16),

$$\begin{aligned} \left(\sum_{x \bmod p} \left| Z_{n,\mu}^N(x) - \frac{1}{p} \right| \right)^2 &\leq \sum_{\substack{\xi \in \mathbb{Z} \\ 0 < |\xi| < \frac{p}{2}}} \left| \hat{Z}_{n,\mu}^N \left(\frac{\xi}{p} \right) \right|^2 \\ &\ll \sum_{0 < |\xi| < \frac{p}{2}} \exp \left(-Cc |\xi|^{\frac{2}{n-1}} \right) \\ &\ll \exp(-Cc). \end{aligned}$$

□

5.1. Proof of Proposition 15, $n = 4$. We first prove Proposition 15 in the case $n = 4$, which gives a good overview of the general argument. The general case introduces a further technical difficulty which is addressed the section that follows.

As for the proof of Theorem 1, the proof of Proposition 15 uses word rearrangement through a group action to introduce smoothing relative to a fixed word in the generators. Let C_2 be the group of two elements. For $k \geq 1$,

$$C_2^2 = \langle \tau_1, \tau_2 : \tau_1^2 = \tau_2^2 = \text{id}, \tau_1 + \tau_2 = \tau_2 + \tau_1 \rangle$$

acts on blocks of indices of length $4k$ with the first factor determining the relative order of the first two blocks of length k , and the second factor determining the relative order of the first two blocks of length k and the second two blocks of length k . Thus if x_1, x_2, x_3, x_4 each represent a block of length k , the group action is given by

$$\begin{aligned} \text{id} \cdot \underline{x} &= x_1 x_2 x_3 x_4 \\ \tau_1 \cdot \underline{x} &= x_2 x_1 x_3 x_4 \\ \tau_2 \cdot \underline{x} &= x_3 x_4 x_1 x_2 \\ \tau_1 \tau_2 \cdot \underline{x} &= x_3 x_4 x_2 x_1. \end{aligned}$$

As in the local limit theorem, abbreviate $\mu^{\otimes N}$ with \mathbb{U}_N . Let $S = \text{supp } \mu$ and $W_N = S^N$. Write

$$Z_{4,\mu}^N = \mathbf{E}_{\mathbb{U}_N} \left[\delta_{Z_4^N(w)} \right].$$

Given a parameter $k \geq 1$, set $N' = N'(k) = \lfloor \frac{N}{4k} \rfloor$. Let $G_k = (C_2^2)^{N'}$ act on W_N with the j th factor acting on the four contiguous blocks of

length k with last index $4jk$. For fixed \underline{w} set

$$Z_k(\underline{w}) = \mathbf{E}_{\underline{\tau} \in G_k} \left[\delta_{Z_4^N(\underline{\tau} \cdot \underline{w})} \right].$$

For each $1 \leq k \leq \frac{N}{4}$,

$$Z_{4,\mu}^N = \mathbf{E}_{\mathbb{U}_N} [Z_k(\underline{w})].$$

The key difference which distinguishes the case $n = 4$ from the case of larger n is that for $n = 4$, the actions of the separate factors in G_k are independent, in the sense that the characteristic function

$$\chi_k(\xi, \underline{w}) = \mathbf{E}_{\underline{\tau} \in G_k} [e_{-\xi}(Z_k(\underline{\tau} \cdot \underline{w}))]$$

factors through the product structure of the G_k action. To check this, let $1 \leq j \leq N'$ and write

$$\begin{aligned} \underline{w}_{j,k,\text{pre}}^{(1)} &= \sum_{1 \leq i \leq 4(j-1)k} w_i^{(1)}, & \underline{w}_{j,k,\text{post}}^{(3)} &= \sum_{4jk+1 \leq i \leq N} w_i^{(3)} \\ \forall 1 \leq i \leq 4, \forall 1 \leq \ell \leq 3, & \underline{w}_{j,k,i}^{(\ell)} &= \sum_{(4j+i-5)k < t \leq (4j+i-4)k} w_t^{(\ell)}. \end{aligned}$$

Thus, setting apart the j th block of size $4k$, $\underline{w}_{\text{pre}}^{(1)}$ is the sum of the first coordinates before the block, $\underline{w}_{\text{post}}^{(3)}$ is the sum of the third coordinates after the block, and $\underline{w}_i^{(\ell)}$ is the sum of the ℓ -coordinates in the i th off-set block of length k .

Given $\underline{\tau} \in G_k$, let $(\hat{\tau}_j, \tau')$ denote $\underline{\tau}$ with $\tau' \in C_2^2$ substituted for τ_j in the j th position. Suppressing the j, k subscripts, for fixed \underline{w} ,

$$\begin{aligned} \Delta_{j,1} &= Z_4^N((\hat{\tau}_j, \tau_1) \cdot \underline{w}) - Z_4^N((\hat{\tau}_j, \text{id}) \cdot \underline{w}) \\ &= \left(\underline{w}_1^{(3)} \underline{w}_2^{(2)} - \underline{w}_1^{(2)} \underline{w}_2^{(3)} \right) \underline{w}_{\text{pre}}^{(1)} \\ &\quad + \left(\underline{w}_1^{(2)} \underline{w}_2^{(1)} - \underline{w}_1^{(1)} \underline{w}_2^{(2)} \right) \left(\underline{w}_3^{(3)} + \underline{w}_4^{(3)} + \underline{w}_{\text{post}}^{(3)} \right) \\ \Delta_{j,2} &= Z_4^N((\hat{\tau}_j, \tau_2) \cdot \underline{w}) - Z_4^N((\hat{\tau}_j, \text{id}) \cdot \underline{w}) \\ &= \left(\left(\underline{w}_3^{(2)} + \underline{w}_4^{(2)} \right) \left(\underline{w}_1^{(3)} + \underline{w}_2^{(3)} \right) - \left(\underline{w}_3^{(3)} + \underline{w}_4^{(3)} \right) \left(\underline{w}_1^{(2)} + \underline{w}_2^{(2)} \right) \right) \underline{w}_{\text{pre}}^{(1)} \\ &\quad + \left(\left(\underline{w}_3^{(1)} + \underline{w}_4^{(1)} \right) \left(\underline{w}_1^{(2)} + \underline{w}_2^{(2)} \right) - \left(\underline{w}_3^{(2)} + \underline{w}_4^{(2)} \right) \left(\underline{w}_1^{(1)} + \underline{w}_2^{(1)} \right) \right) \underline{w}_{\text{post}}^{(3)} \\ &\quad + \underline{w}_3^{(1)} \underline{w}_4^{(2)} \left(\underline{w}_1^{(3)} + \underline{w}_2^{(3)} \right) + \left(\underline{w}_3^{(1)} + \underline{w}_4^{(1)} \right) \underline{w}_1^{(2)} \underline{w}_2^{(3)} \\ &\quad - \underline{w}_1^{(1)} \underline{w}_2^{(2)} \left(\underline{w}_3^{(3)} + \underline{w}_4^{(3)} \right) - \left(\underline{w}_1^{(1)} + \underline{w}_2^{(1)} \right) \underline{w}_3^{(2)} \underline{w}_4^{(3)} \\ \Delta_{j,12} &= Z_4^N((\hat{\tau}_j, \tau_1 \tau_2) \cdot \underline{w}) - Z_4^N((\hat{\tau}_j, \text{id}) \cdot \underline{w}) \\ &= \Delta_{j,1} + \Delta_{j,2} + \delta_j \\ \delta_j &:= \left(\underline{w}_1^{(1)} \underline{w}_2^{(2)} - \underline{w}_1^{(2)} \underline{w}_2^{(1)} \right) \left(\underline{w}_3^{(3)} + \underline{w}_4^{(3)} \right) \\ &\quad + \left(\underline{w}_1^{(3)} \underline{w}_2^{(2)} - \underline{w}_1^{(2)} \underline{w}_2^{(3)} \right) \left(\underline{w}_3^{(1)} + \underline{w}_4^{(1)} \right). \end{aligned}$$

Since $\Delta_{j1}, \Delta_{j2}, \Delta_{j12}$ are each invariant under the action of the factors of G_k outside the j th position,

$$\begin{aligned} \chi_k(\xi, \underline{w}) &= e_{-\xi}(Z_4^N(\underline{w})) \\ &\times \prod_{j=1}^{N'} \left[\frac{1}{4} (1 + e_{-\xi}(\Delta_{j,1}) + e_{-\xi}(\Delta_{j,2}) + e_{-\xi}((\Delta_{j,1} + \Delta_{j,2} + \delta_j))) \right]. \end{aligned}$$

In particular, for some $C > 0$,

$$(17) \quad |\chi_k(\xi, \underline{w})| \leq \prod_{j=1}^{N'} \exp(-C \|\xi \delta_j\|_{\mathbb{R}/\mathbb{Z}}^2).$$

Assume without loss that $|\xi| \geq N^{-\frac{3}{2}}$. Let N_0 be minimal such that e_1, e_2, e_3 may each be formed by words of length at most N_0 on vectors from $\text{supp } \mu$. Let $k = \max\left(N_0, \left\lfloor \frac{1}{|\xi|^{\frac{2}{3}}} \right\rfloor\right)$. Note that on average over \underline{w} , the $\{\delta_j\}_{1 \leq j \leq N'}$ are i.i.d., that δ_j has limiting distribution as $k \rightarrow \infty$ at scale $\frac{1}{|\xi|}$, that the limiting distribution has a density on \mathbb{R} which is positive, and thus if $\epsilon(\mu)$ is sufficiently small that the event $E_j = \left\{ \|\xi \delta_j\|_{\mathbb{R}/\mathbb{Z}} \geq \epsilon(\mu) \right\}$ occurs with probability at least $p_0(\mu) > 0$ uniformly in ξ . The estimate, for some $c > 0$,

$$\left| \hat{Z}_{4,\mu}^N \right| \leq \mathbf{E}_{\mathbb{U}_N} [|\chi_k(\xi, \underline{w})|] \leq \exp\left(-cN|\xi|^{\frac{3}{2}}\right)$$

follows, see the proof of Lemma 12.

5.2. Proof of Proposition 15, general case. Consider $n \geq 5$. Let C_2^{n-2} act on blocks of vectors of length $k2^{n-2}$ with the j th factor from C_2^{n-2} , $j \geq 1$ switching the relative order of the first $k2^{j-1}$ and second $k2^{j-1}$ indices. Thus, for instance, in case $n = 5$, if each of x_1, \dots, x_8 represents a block of k consecutive indices and $\underline{x} = x_1x_2x_3x_4x_5x_6x_7x_8$,

$$\begin{aligned} \tau_2 \underline{x} &= x_3x_4x_1x_2x_5x_6x_7x_8 \\ \tau_1 \tau_3 \underline{x} &= \tau_3 \tau_1 \underline{x} = x_5x_6x_7x_8x_2x_1x_3x_4 \\ \tau_1 \tau_2 \tau_3 \underline{x} &= x_5x_6x_7x_8x_3x_4x_2x_1. \end{aligned}$$

For $k \geq 1$ again set $N' = \lfloor \frac{N}{k2^{n-2}} \rfloor$ and let $G_k = (C_2^{n-2})^{N'}$. G_k acts on $W_N = (\text{supp } \mu)^N$ with, for $j \geq 1$, the j th factor of G_k acting on the contiguous subsequence of indices of length $k2^{n-2}$ ending at $jk2^{n-2}$. For fixed k and fixed $\underline{w} \in W_N$, let

$$Z_k(\underline{w}) = \mathbf{E}_{\mathbb{T} \in G_k} [\delta_{Z_n^N(\mathbb{T} \cdot \underline{w})}].$$

Continue to abbreviate $\mathbb{U}_N = \mu^{\otimes N}$. For any k ,

$$Z_{n,\mu}^N = \mathbf{E}_{\mathbb{U}_N} [Z_k(\underline{w})].$$

The action of G_k on \underline{w} has a dual action on a linear space of dual n -tensors. Let \mathcal{J}_N be the collection of a multi-indices $\underline{i} = (i_1, i_2, \dots, i_{n-1})$

satisfying $1 \leq i_1 < i_2 < \dots < i_{n-1} \leq N$. Given $\underline{i} \in \mathcal{I}_N$ and $k \geq 1$, let $\mathfrak{S}_{\underline{i},k} \subset \mathfrak{S}_{n-1}$ be the subset of permutations $\mathfrak{S}_{\underline{i},k} = \{\sigma_{\underline{\tau},\underline{i}} : \underline{\tau} \in G_k\}$ where

$$\forall 1 \leq j \leq n-1, \quad \sigma_{\underline{\tau},\underline{i}}(j) = \#\{1 \leq k \leq n-1 : \underline{\tau}(i_k) \leq \underline{\tau}(i_j)\}.$$

That is, $\sigma_{\underline{\tau},\underline{i}}(j)$ is the relative position of $\underline{\tau} \cdot i_j$ when $\underline{\tau} \cdot \underline{i}$ is sorted to be in increasing order. Put another way, suppose $\underline{\tau}$ maps $i_1 < \dots < i_{n-1}$ to $j_1 < \dots < j_{n-1}$ in some order (and vice versa, τ is an involution) and calculate

$$\begin{aligned} e_{j_1}^{(1)} \otimes \dots \otimes e_{j_{n-1}}^{(n-1)}(\underline{\tau} \cdot \underline{w}) &= e_{\underline{\tau} \cdot j_1}^{(1)} \otimes \dots \otimes e_{\underline{\tau} \cdot j_{n-1}}^{(n-1)}(\underline{w}) \\ &= e_{i_{\sigma^{-1}(1)}}^{(1)} \otimes \dots \otimes e_{i_{\sigma^{-1}(n-1)}}^{(n-1)}(\underline{w}) \\ &= e_{i_1}^{(\sigma(1))} \otimes \dots \otimes e_{i_{n-1}}^{(\sigma(n-1))}(\underline{w}), \end{aligned}$$

where $\sigma_{\underline{\tau},\underline{i}}$ is abbreviated σ .

Let

$$X_{N,k} = \left\{ e_{i_1}^{(\sigma(1))} \otimes \dots \otimes e_{i_{n-1}}^{(\sigma(n-1))} : \underline{i} \in \mathcal{I}_N, \sigma \in \mathfrak{S}_{\underline{i},k} \right\}.$$

The action of $\underline{\tau} \in G_k$ is defined on a representative set within $X_{N,k}$ by, for each $\underline{i} \in \mathcal{I}_N$,

$$(18) \quad \underline{\tau} \cdot \left(e_{i_1}^{(1)} \otimes \dots \otimes e_{i_{n-1}}^{(n-1)} \right) = e_{i_1}^{(\sigma_{\underline{\tau},\underline{i}}(1))} \otimes \dots \otimes e_{i_{n-1}}^{(\sigma_{\underline{\tau},\underline{i}}(n-1))}.$$

The following lemma justifies that this definition extends to a unique group action of G_k on all of $X_{N,k}$.

Lemma 16. *Let $\underline{\tau}, \underline{\tau}' \in G_k$ and $\underline{i} \in \mathcal{I}_N$ satisfy $\sigma_{\underline{\tau},\underline{i}} = \sigma_{\underline{\tau}',\underline{i}}$. Then for any $\underline{\tau}'' \in G_k$, $\sigma_{\underline{\tau}+\underline{\tau}'',\underline{i}} = \sigma_{\underline{\tau}'+\underline{\tau}'',\underline{i}}$. In particular, (18) extends to a unique group action on $X_{N,k}$.*

Proof. This follows, since, for any $1 \leq i < j \leq N$ there is at most one factor of $G = C_2^{m-2}$ in G_k , and one index ℓ , $1 \leq \ell \leq n-2$ of G which exchanges the order of i and j . To define the group action in general, given $\underline{\tau} \in G_k$, $\underline{i} \in \mathcal{I}_N$ and $\sigma \in \mathfrak{S}_{\underline{i},k}$ choose any $\underline{\tau}_0$ such that $\sigma = \sigma_{\underline{\tau}_0,\underline{i}}$. Let $\sigma' = \sigma_{\underline{\tau}_0+\underline{\tau},\underline{i}}$. Then

$$\underline{\tau} \cdot e_{i_1}^{(\sigma(1))} \otimes \dots \otimes e_{i_{n-1}}^{(\sigma(n-1))} = e_{i_1}^{(\sigma'(1))} \otimes \dots \otimes e_{i_{n-1}}^{(\sigma'(n-1))}.$$

The definition is clearly unique, since (18) surjects on $X_{N,k}$. \square

The actions of \mathfrak{S}_N on W_N and on X_N , although not adjoint, are compatible on Z_n^N , in the sense that for any $\underline{\tau}$,

$$(\underline{\tau} \cdot Z_n^N)(\underline{w}) = Z_n^N(\underline{\tau} \cdot \underline{w})$$

so that

$$Z_k(\underline{w}) = \mathbf{E}_{\underline{\tau} \in G_k} [\delta_{(\underline{\tau} \cdot Z_n^N)(\underline{w})}].$$

Note that when $n \geq 5$, although G_k is a product group, the separate factors τ_i do not act independently in $\underline{\tau} \cdot Z_n^N$ in the sense that the characteristic function

$$\chi_k(\xi, \underline{w}) = \mathbf{E}_{\underline{\tau} \in G_k} [e_{-\xi} (Z_n^N(\underline{\tau} \cdot \underline{w}))]$$

need not factor as a product. A pleasant feature of the general case is that this difficulty is rectified by estimating instead of $\chi_k(\xi, \underline{w})$, a function $F_k(\xi, \underline{w})$ which is the result of applying the Gowers-Cauchy-Schwarz inequality to $\chi_k(\xi, \underline{w})$. To describe this, write $G_k = (C_2^{n-2})^{N'} = (C_2^{N'})^{n-2}$, and thus

$$\mathbf{E}_{\underline{\tau} \in G_k} [f(\underline{\tau})] = \mathbf{E}_{\tau_1 \in C_2^{N'}} \cdots \mathbf{E}_{\tau_{n-2} \in C_2^{N'}} [f(\tau_1, \dots, \tau_{n-2})].$$

Then, setting apart one expectation at a time and applying Cauchy-Schwarz,

$$\begin{aligned} & |\chi_k(\xi, \underline{w})|^{2^{n-2}} \\ & \leq \mathbf{E}_{\tau_1, \tau'_1 \in C_2^{N'}} \cdots \mathbf{E}_{\tau_{n-2}, \tau'_{n-2} \in C_2^{N'}} \left[e_{-\xi} \left(\sum_{S \subset [n-2]} (-1)^{n-2-|S|} \underline{\tau}_S \cdot \underline{w} \right) \right] \\ & = \mathbf{E}_{\underline{\tau}, \underline{\tau}' \in G_k} \left[e_{-\xi} \left(\sum_{S \subset [n-2]} (-1)^{n-2-|S|} \underline{\tau}_S \cdot \underline{w} \right) \right] \\ & =: F_k(\xi, \underline{w}), \end{aligned}$$

where

$$\underline{\tau}_S = (\tau_{S,1}, \dots, \tau_{S,n-2}), \quad \tau_{S,i} = \begin{cases} \tau_i & i \in S \\ \tau'_i & i \notin S \end{cases}$$

Lemma 17. $F_k(\xi, \underline{w})$ factors as the product

$$F_k(\xi, \underline{w}) = \prod_{j=1}^{N'} \left(1 - \frac{1}{2^{n-2}} + \frac{F_{k,j}(\xi, \underline{w})}{2^{n-2}} \right)$$

where $F_{k,j}(\xi, \underline{w})$ is a function of $\underline{w}_{j,k} = (\omega_1, \dots, \omega_{2^{n-2}})$ with the

$$\omega_i = \sum_{(2^{n-2}(j-1)+i-1)k < \ell \leq (2^{n-2}(j-1)+i)k} w_\ell$$

the sum of consecutive blocks of length k in \underline{w} . Identify C_2^{n-2} with $\{0, 1\}^{n-2}$ and write $|\tau| = \sum_{i=1}^{n-2} \mathbf{1}(\tau_i \neq 0)$. Then

$$F_{k,j}(\xi, \underline{w}) = \mathbf{E}_{\tau \in C_2^{n-2}} \left[e_{-\xi} \left(\sum_{\tau' \in C_2^{n-2}} (-1)^{|\tau'|} Z_n^{2^{n-2}}((\tau + \tau') \cdot \underline{w}_{j,k}) \right) \right]$$

with the action of C_2^{n-2} on blocks of size 1 in $\underline{w}_{j,k}$.

Proof. Consider for fixed $\underline{\tau}, \underline{\tau}' \in G_k$ the sum

$$Z_n^N(\underline{\tau}, \underline{\tau}')(\underline{w}) = \sum_{S \subset [n-2]} (-1)^{n-2-|S|} \underline{\tau}_S \cdot Z_n^N(\underline{w}).$$

After replacing \underline{w} with $\underline{\tau}' \underline{w}$ and $\underline{\tau}$ with $\underline{\tau} + \underline{\tau}'$ it suffices to consider $\underline{\tau}' = \text{id}$.

Consider the action of

$$\hat{\tau} = \sum_{S \subset [n-2]} (-1)^{n-2-|S|} \tau_S$$

on a tensor

$$\mathbf{e} = e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \cdots \otimes e_{i_{n-1}}^{(n-1)}, \quad 1 \leq i_1 < i_2 < \cdots < i_{n-1} \leq N$$

appearing in Z_n^N . Let $G = C_2^{n-2}$ identified with subsets S of $[n-2]$, let $G^0 = \text{stab}(\underline{i}) \leq C_2^{n-2}$ be the subgroup consisting of S for which $\underline{\tau}_S \cdot \mathbf{e} = \mathbf{e}$, and let $G^1 = C_2^{n-2}/G^0$. By the group action property, for all $x \in G^0$, for all $y \in G^1$, $\underline{\tau}_{x+y} \mathbf{e} = \underline{\tau}_y \mathbf{e}$ so that when $G^0 \neq \{1\}$, $\hat{\tau} \cdot \mathbf{e} = 0$.

A necessary and sufficient condition for $G^0 = \{1\}$ is that, for some $1 \leq j \leq N'$,

$$2^{n-2}(j-1)k < i_1 \leq 2^{n-2}(j-1)k + k$$

$$\forall 1 < \ell \leq n-1 \quad 2^{n-2}(j-1)k + 2^{\ell-2}k < i_\ell \leq 2^{n-2}(j-1)k + 2^{\ell-1}k,$$

and $\tau_j = \mathbf{1}_{n-2} \in C_2^{n-2}$. In words, the indices must all belong to a common block of length $2^{n-2}k$ acted on by a single factor from G_k , within this block, the first $2^{\ell-1}k$ elements must contain i_ℓ and the second $2^{\ell-1}k$ must contain $i_{\ell+1}$ for $\ell = 1, 2, \dots, n-2$, and the factor τ_j acting on the block must be the element $\mathbf{1}_{n-2}$ of the hypercube C_2^{n-2} .

The product formula given summarizes this condition. □

The remainder of the proof of the general case of Proposition 15 now follows essentially as in the case $n = 4$.

APPENDIX A. THE CHARACTERISTIC FUNCTION OF A GAUSSIAN MEASURE ON THE HEISENBERG GROUP

This section gives the proof of Theorem 9, which gives a rate of convergence to Gaussian measure on the Heisenberg group.

Recall that

$$I(\alpha, \xi; N) = \int_{(\mathbb{R}^2)^N} e_{-\alpha} \left(\frac{\underline{x}}{\sqrt{N}} \right) e_{-\xi} \left(\frac{H^*(\underline{x})}{N} \right) d\nu_2^{\otimes N}(\underline{x})$$

where $\nu_2(x) = \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right)$.

First consider the case $\alpha = 0$. Integrate away $\underline{x}^{(1)}$ to obtain,

$$I(0, \xi; N) = \frac{1}{(2\pi)^{\frac{N}{2}}} \int_{\mathbb{R}^N} \exp\left(-\frac{1}{2} \underline{y}^t \left((1 - \xi_0^2) I_N + \xi_0^2 H\right) \underline{y}\right) d\underline{y}$$

where

$$\xi_0 = \frac{\pi\xi}{N}, \quad H_{i,j} = N - 2|i - j|;$$

this follows from

$$\underline{y}^t (H - I_N) \underline{y} = \sum_{i=1}^N \left(\sum_{j \neq i} (-1)^{\delta(j < i)} y_j \right)^2.$$

Thus,

$$I(0, \xi; N) = \frac{1}{\sqrt{\det((1 - \xi_0^2) I_N + \xi_0^2 H)}}.$$

Let

$$U_- = I_N - \sum_{i=1}^{N-1} e_i \otimes e_{i+1}.$$

Then

$$\begin{aligned} & U_-^t ((1 - \xi_0^2) I_N + \xi_0^2 H) U_- \\ &= (1 + \xi_0^2) \left[2I_N - \frac{1 - \xi_0^2}{1 + \xi_0^2} \sum_{i=1}^{N-1} (e_i \otimes e_{i+1} + e_{i+1} \otimes e_i) \right. \\ & \quad \left. - \frac{2\xi_0^2}{1 + \xi_0^2} \sum_{i=1}^N (e_1 \otimes e_i + e_i \otimes e_1) + \frac{(N+1)\xi_0^2 - 1}{1 + \xi_0^2} e_1 \otimes e_1 \right]. \end{aligned}$$

Set $\eta = \frac{1 - \xi_0^2}{1 + \xi_0^2}$ and define sequences

$$\begin{aligned} \varepsilon_1 &= 2, & \forall i \geq 1, \varepsilon_{i+1} &= 2 - \frac{\eta^2}{\varepsilon_i} \\ \pi_0 &= 1, & \forall i \geq 1, \pi_i &= \prod_{j=1}^i \varepsilon_j \\ \delta_1 &= 1, & \forall i \geq 1, \delta_{i+1} &= 1 + \frac{\eta\delta_i}{\varepsilon_i} \end{aligned}$$

These parameters have the following behavior with proof postponed until the end of this section.

Lemma 18. For $\xi \in \left(0, N^{\frac{1}{2}}\right]$ the following asymptotics hold

$$\begin{aligned}\pi_N &= N \frac{\sinh(2\pi\xi)}{2\pi\xi} \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right) \\ \varepsilon_N &= 1 + \frac{2\pi\xi}{N} \coth(2\pi\xi) \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right) \\ \delta_N &= \frac{N \tanh \pi\xi}{2\pi\xi} \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right) \\ \sum_{j=1}^{N-1} \frac{\delta_j^2}{\varepsilon_j} &= \frac{N^3}{8\pi^3\xi^3} [2\pi\xi - 2 \tanh \pi\xi] \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right).\end{aligned}$$

Set

$$L_\varepsilon = I_N + \eta \sum_{i=1}^{N-1} \frac{e_{i+1} \otimes e_i}{\varepsilon_{N-i+1}}, \quad D_\varepsilon = \frac{1}{(1+\xi_0^2)} \sum_{i=1}^N \frac{e_i \otimes e_i}{\varepsilon_{N+1-i}}.$$

Then

$$D_\varepsilon^{\frac{1}{2}} L_\varepsilon^t U_-^t \left((1 - \xi_0^2) I_N + \xi_0^2 H \right) U_- L_\varepsilon D_\varepsilon^{\frac{1}{2}} = I_N + P$$

where P is the rank two symmetric matrix

$$P = \frac{-2\xi_0^2}{1+\xi_0^2} \sum_{i=1}^N \frac{\delta_{N+1-i}}{\sqrt{\varepsilon_N \varepsilon_{N-i+1}}} (e_1 \otimes e_i + e_i \otimes e_1) + \frac{(N+1)\xi_0^2 - 1}{\varepsilon_N(1+\xi_0^2)} e_1 \otimes e_1.$$

Then, for some orthogonal matrix O , and $\lambda_+ \geq \lambda_-$,

$$O^t(I_N + P)O = (\lambda_+ e_1 \otimes e_1 + \lambda_- e_2 \otimes e_2) \oplus I_{N-2}.$$

By direct calculation (expand by the top row),

$$\begin{aligned}(19) \quad \det(I_N + P) &= \lambda_+ \lambda_- = \left(1 - \frac{1}{\varepsilon_N(1+\xi_0^2)}\right) + \frac{(N+1)\xi_0^2}{\varepsilon_N(1+\xi_0^2)} \\ &\quad - \frac{4\xi_0^2}{1+\xi_0^2} \frac{\delta_N}{\varepsilon_N} - \frac{4\xi_0^4}{(1+\xi_0^2)^2 \varepsilon_N} \sum_{j=1}^{N-1} \frac{\delta_j^2}{\varepsilon_j} \\ &= \frac{\pi\xi \coth \pi\xi}{N} \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right).\end{aligned}$$

Since

$$(20) \quad \det(D_\varepsilon)^{-1} = (1+\xi_0^2)^N \pi_N = \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right) \frac{N \sinh 2\pi\xi}{2\pi\xi},$$

$$\det((1 - \xi_0^2) I_N + \xi_0^2 H) = (\cosh \pi\xi)^2 \left(1 + O\left(\frac{1+\xi^2}{N}\right)\right).$$

Now consider the general case in which $\alpha \neq 0$. Treat \underline{x} as N vectors in \mathbb{R}^2 . When $\text{SO}_2(\mathbb{R})$ acts diagonally on $(\mathbb{R}^2)^N$ rotating each x_i

simultaneously, H^* and the Gaussian density are preserved. Thus, $I(\alpha, \xi; N) = I((0, \|\alpha\|)^t, \xi; N)$. Calculate

$$[1, 1, \dots, 1]U_-L_\epsilon D_\epsilon^{\frac{1}{2}} = \frac{e_1}{\sqrt{\epsilon_N(1 + \xi_0^2)}}.$$

It follows that after making the change of coordinates $\underline{y}' =: U_1 L_\epsilon D_\epsilon^{\frac{1}{2}} \underline{y}$ the phase has magnitude $\frac{2\pi\|\alpha\|}{\sqrt{N\epsilon_N(1+\xi_0^2)}}$ and is now in the e_1 direction. Let v_+, v_- be unit vectors generating the eigenspaces λ_+, λ_- respectively. Since e_1 lies in the span of v_+, v_- it follows

$$I(\alpha, \xi; N) = \exp\left(\frac{-2\pi^2\|\alpha\|^2}{N\epsilon_N(1 + \xi_0^2)}\left(\frac{\langle v_+, e_1 \rangle^2}{\lambda_+} + \frac{\langle v_-, e_1 \rangle^2}{\lambda_-}\right)\right) I(0, \xi).$$

Calculate

$$(21) \quad T = \lambda_+ + \lambda_- = 2 + e_1^t P e_1 = 1 + O\left(\frac{\xi^2}{N}\right)$$

so that

$$(\lambda_+, \lambda_-) = \left(1 + O\left(\frac{1 + \xi^2}{N}\right)\right) \left(1, \frac{\pi\xi \coth \pi\xi}{N}\right).$$

Also,

$$\begin{aligned} \langle v_+, e_1 \rangle^2 + \langle v_-, e_1 \rangle^2 &= 1 \\ \lambda_+ \langle v_+, e_1 \rangle^2 + \lambda_- \langle v_-, e_1 \rangle^2 &= 1 + e_1^t P e_1 = O\left(\frac{1 + \xi^2}{N}\right) \end{aligned}$$

so that

$$\langle v_+, e_1 \rangle^2 = O\left(\frac{1 + \xi^2}{N}\right), \quad \langle v_-, e_1 \rangle^2 = 1 + O\left(\frac{1 + \xi^2}{N}\right).$$

It follows that

$$\frac{\langle v_+, e_1 \rangle^2}{\lambda_+} + \frac{\langle v_-, e_1 \rangle^2}{\lambda_-} = \frac{N}{\pi\xi \coth \pi\xi} \left(1 + O\left(\frac{1 + \xi^2}{N}\right)\right).$$

In particular

$$I(\alpha, \xi; N) = \frac{\exp\left(\frac{-2\pi\|\alpha\|^2}{\xi \coth \pi\xi}\right)}{\cosh \pi\xi} \left(1 + O\left(\frac{(1 + \|\alpha\|^2)(1 + \xi^2)}{N}\right)\right)$$

Proof of Lemma 18. Recall $\eta = \frac{1 - \xi_0^2}{1 + \xi_0^2}$. π_n satisfies the recurrence

$$\pi_n = 2\pi_{n-1} - \eta^2 \pi_{n-2}, \quad \pi_0 = 1, \pi_1 = 2.$$

The following closed forms hold,

$$\begin{aligned}\pi_n &= \frac{(1 + \xi_0)^{2n+2} - (1 - \xi_0)^{2n+2}}{4\xi_0(1 + \xi_0^2)^n} \\ \varepsilon_n &= 1 + \frac{2\xi_0}{1 + \xi_0^2} \frac{(1 + \xi_0)^{2n} + (1 - \xi_0)^{2n}}{(1 + \xi_0)^{2n} - (1 - \xi_0)^{2n}} \\ \delta_n &= \frac{1}{2\xi_0} \frac{\left(\frac{1+\xi_0}{1-\xi_0}\right)^n + \left(\frac{1-\xi_0}{1+\xi_0}\right)^n - 2}{\left(\frac{1+\xi_0}{1-\xi_0}\right)^n - \left(\frac{1-\xi_0}{1+\xi_0}\right)^n} + \frac{1}{2}\end{aligned}$$

The formula for π_n is immediate from the recurrence relation, since

$$\frac{(1 + \xi_0)^2}{1 + \xi_0^2}, \quad \frac{(1 - \xi_0)^2}{1 + \xi_0^2}$$

are the two roots of $x^2 - 2x + \eta^2 = 0$. The formula for ε_n follows from $\varepsilon_n = \frac{\pi_n}{\pi_{n-1}}$. The formula for δ_n is obtained on summing the geometric series

$$\delta_n = \frac{\eta^{n-1}}{\pi_{n-1}} \sum_{j=0}^{n-1} \frac{\pi_j}{\eta^j},$$

and use

$$\begin{aligned}\frac{\pi_n}{\eta^n} &= \frac{(1 + \xi_0)^{2n+2} - (1 - \xi_0)^{2n+2}}{4\xi_0(1 - \xi_0^2)^n} \\ &= \frac{1 - \xi_0^2}{4\xi_0} \left[\left(\frac{1 + \xi_0}{1 - \xi_0}\right)^{n+1} - \left(\frac{1 - \xi_0}{1 + \xi_0}\right)^{n+1} \right].\end{aligned}$$

The claimed asymptotics for π, ε, δ are straightforward.

Using the estimates for ε and δ yields

$$\begin{aligned}\frac{\delta_j^2}{\varepsilon_j} &= \left(1 + O\left(\frac{1}{j} + \frac{\xi^2}{N}\right)\right) \frac{N^2}{\xi^2} \left(\frac{\cosh\left(\frac{j\xi}{N}\right) - 1}{\sinh\left(\frac{j\xi}{N}\right)}\right)^2 \\ &= \left(1 + O\left(\frac{1}{j} + \frac{\xi^2}{N}\right)\right) \frac{N^2}{\xi^2} \tanh\left(\frac{j\xi}{2N}\right)^2.\end{aligned}$$

Approximating with a Riemann sum,

$$\sum_{j=1}^{N-1} \frac{\delta_j^2}{\varepsilon_j} = \left(1 + O\left(\frac{\xi^2}{N}\right)\right) \frac{N^3}{\xi^2} \int_0^1 \tanh\left(\frac{t\xi}{2}\right)^2 dt$$

which gives the claimed estimate. \square

REFERENCES

- [1] G. K. Alexopoulos “Random walks on discrete groups of polynomial volume growth.” *Ann. Probab.* **30** (2002), no. 2, 723–801.
- [2] G. K. Alexopoulos “Random walks on discrete groups of polynomial volume growth.” *Ann. Probab.* **30** (2002), no. 2, 723–801.

- [3] P. Baldi and L. Caramellino. “Large and moderate deviations for random walks on nilpotent groups.” *J. Theoret. Probab.* **12** (1999), no. 3, 779–809. MR1702883 (2001b:60013)
- [4] Balog, A. “On the distribution of integers having no large prime factors.” *Journées Arithmétiques*, Besançon, Astérisque 147/148 (1985): 27–31.
- [5] E. F. Breuillard, *Equidistribution of random walks on nilpotent Lie groups and homogeneous spaces*, ProQuest LLC, Ann Arbor, MI, 2004.
- [6] Breuillard, Emmanuel. “Local limit theorems and equidistribution of random walks on the Heisenberg group.” *Geometric and functional analysis GAFA* 15.1 (2005): 35–82.
- [7] Breuillard, Emmanuel. “Equidistribution of dense subgroups on nilpotent Lie groups.” *Ergodic Theory Dynam. Systems* 30 (2010), no. 1, 131150.
- [8] Daniel Bump, Persi Diaconis, Angela Hicks, Laurent Miclo, and Harold Widom. An exercise (?) in Fourier analysis on the Heisenberg group. arXiv:1502.04160, 2015.
- [9] P. Crépél and A. Raugi. “Théorème central limite sur les groupes nilpotents.” *Ann. Inst. H. Poincaré* sec B, Prob. and Stat., vol XIV, 2. (1978): 145–164.
- [10] Coulhon, Saloff-Coste, and Varopoulos. *Analysis and geometry on groups*. Cambridge tracts on mathematics, Cambridge University Press, 1992.
- [11] Chung, Fan and Linyuan Lu. “Concentration inequalities and martingale inequalities: a survey.” *Internet mathematics* 3.1 (2006): 79–127.
- [12] P. Diaconis. “Threads through group theory.” In *Character theory of finite groups*, 33–47, Contemp. Math., 524, Amer. Math. Soc., Providence, RI (2010).
- [13] P. Diaconis and L. Saloff-Coste. “Moderate growth and random walk on finite groups.” *Geom. Funct. Anal.* **4** (1994), no. 1, 1–36.
- [14] P. Diaconis and L. Saloff-Coste. “An application of Harnack inequalities to random walk on nilpotent quotients.” *J. Fourier Anal. Appl.* **1995**, Special Issue, 189–207.
- [15] P. Diaconis and L. Saloff-Coste. “Nash inequalities for finite Markov chains.” *J. Theoret. Probab.* **9** (1996), no. 2, 459–510.
- [16] Lawler, Gregory F., and Vlada Limic. *Random walk: a modern introduction*. Vol. 123. Cambridge University Press, 2010.
- [17] Gaveau, Bernard. “Principe de moindre action, propagation de la chaleur et estimés sous elliptiques sur certains groupes nilpotents.” *Acta mathematica* 139.1 (1977): 95–153.
- [18] Green, Ben; Tao, Terence. “The quantitative behaviour of polynomial orbits on nilmanifolds.” *Ann. of Math.* (2) 175 (2012), no. 2, 465–540.
- [19] Hulanicki, Andrzej. “The distribution of energy in the Brownian motion in the Gaussian field and analytic-hypoellipticity of certain subelliptic operators on the Heisenberg group.” *Studia Mathematica* 2.56 (1976): 165–173.
- [20] S. Ishiwata. “A central limit theorem on a covering graph with a transformation group of polynomial growth.” *J. Math. Soc. Japan* **55** (2003), no. 3, 837–853.
- [21] S. Ishiwata. “A central limit theorem on modified graphs of nilpotent covering graphs.” In *Spectral analysis in geometry and number theory*, 59–72, Contemp. Math., 484, Amer. Math. Soc., Providence, RI.
- [22] Y. Peres and A. Sly. “Mixing of the upper triangular matrix walk.” *Probab. Theory Related Fields* **156** (2013), no. 3–4, 581–591.
- [23] Raugi, A. “Thorme de la limite centrale sur les groupes nilpotents.” *Probability Theory and Related Fields* 43.2 (1978): 149–172.
- [24] L. Saloff-Coste, “Probability on groups: random walks and invariant diffusions.” *Notices Amer. Math. Soc.* **48** (2001), no. 9, 968–977.

- [25] D. W. Stroock and S. R. S. Varadhan. Limit theorems for random walks on Lie groups. *Sankhyā* Ser. A **35** (1973), no. 3, 277–294. MR0517406 (58 #24457)
- [26] V.N. Tutubalin. “Compositions of measures on the simplest nilpotent group.” (Russian) *Teor. Veroyatnost. i Primenen* 9, (1964): 531–539.
- [27] D. Wehn. “Probabilities on Lie groups.” *Proc. Nat. Acad. Sci. U.S.A.* 48 (1962): 791–795.

(Persi Diaconis) DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, 450
SERRA MALL, STANFORD, CA, 94305, USA

E-mail address: `diaconis@math.stanford.edu`

(Bob Hough) DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, 450
SERRA MALL, STANFORD, CA, 94305, USA

Current address: School of Mathematics, Institute of Advanced Study, 1 Einstein
Drive, Princeton, NJ, 08540

E-mail address: `hough@math.ias.edu`